Adaptive Equalisation for Impulsive Noise Environments

Apostolos Theofani Georgiadis

Απόστολος Θεοφάνη Γεωργιάδης

A thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

September 2000

Abstract

This thesis addresses the problem of adaptive channel equalisation in environments where the interfering noise exhibits non–Gaussian behaviour due to impulsive phenomena. The family of α–stable distributions has proved to be a suitable and flexible tool for the modelling of signals with impulsive nature. However, non–Gaussian α–stable signals have infinite variance, and signal processing techniques based on second order moments are meaningless in such environments. In order to exploit the flexibility of the stable family and still take advantage of the existing signal processing tools, a novel framework for the integration of the stable model in a communications context is proposed, based on a finite dynamic range receiver. The performance of traditional signal processing algorithms designed under the Gaussian assumption may degrade seriously in impulsive environments. When this degradation cannot be tolerated, the traditional signal processing methods must be revisited and redesigned taking into account the non–Gaussian noise statistics. In this direction, the optimum feed–forward and decision feedback Bayesian symbol–by–symbol equalisers for stable noise environments are derived. Then, new analytical tools for the evaluation of systems in infinite variance environments are presented. For the centers estimation of the proposed Bayesian equaliser, a unified framework for a family of robust recursive linear estimation techniques is presented and the underlying relationships between them are identified. Furthermore, the direct clustering technique is studied and robust variants of the existing algorithms are proposed. A novel clustering algorithm is also derived based on robust location estimation. The problem of estimating the stable parameters has been addressed in the literature and a variety of algorithms can be found. Some of these algorithms are assessed in terms of efficiency, simplicity and performance and the most suitable is chosen for the equalisation problem. All the building components of an adaptive Bayesian equaliser are then put together and the performance of the equaliser is evaluated experimentally. The simulation results suggest that the proposed adaptive equaliser offers a significant performance benefit compared with a traditional equaliser, designed under the Gaussian assumption. The implementation of the proposed Bayesian equaliser is simple but the computational complexity can be unaffordable. However, this thesis proposes certain approximations which enable the computationally efficient implementation of the optimum equaliser with negligible loss in performance.

Declaration of originality

I hereby declare that the research recorded in this thesis and the thesis itself was composed and originated entirely by myself in the Department of Electronics and Electrical Engineering at The University of Edinburgh.

The software used to perform the simulations was written by myself with the following exceptions:

- The routines used to generate uniform and Gaussian distributed pseudo–random samples were obtained from *Numerical recipes in C* [1].
- The routine for the generation of α–stable distributed pseudo–random samples was written by Dr. Stephen Bates.

Apostolos Georgiadis

to my parents, Theofanis and Chrysi.

στους γονείς μου, Θεοφάνη και Χρυσή.

Acknowledgements

I wish to express my gratitude to my supervisor Prof. Bernard Mulgrew for his guidance and invaluable advice, for sharing his insight, for our numerous discussions and for reviewing the manuscript. Also, my second supervisor Dr. Stephen McLaughlin for his advice and guidance when Prof. Mulgrew was away, and for reviewing the manuscript.

I would like to thank the University of Edinburgh for the financial support of this work.

My sincere thanks to my friends and colleagues in the *Signals and Systems* group for sharing their knowledge, and especially George Vardoulias, Sarat Patra, Stephen Bates and Mark Cowper.

I wish to thank my good friend George Vardoulias for the beautiful time we shared together.

I express my gratitude to Barbara Mitliaga for her endless support and love, for sharing our dreams, and for making our lives so bright.

I would like to express my deepest gratitude to my parents Theofanis and Chrysi for their relentless support, priceless guidance and boundless love.

Contents

List of figures

List of tables

Acronyms and abbreviations

BER	bit–error ratio
bps	bits per second
DCS	digital communication system
DFE	decision feedback equaliser
eq.	equation
fig.	figure
FLOM	fractional lower order moment
i.i.d.	independent identically distributed
ISI	intersymbol interference
LMP	least mean p–norm
LMS	least mean squares
LS	least squares
MAP	maximum a posteriori probabilities
MAPSD	MAP symbol–by–symbol detector
MD	minimum dispersion
MLMP	modified LMP
MLSE	maximum likelihood sequence estimator
MLVA	maximum likelihood Viterbi algorithm
MMPE	minimum mean p–norm of the error
MMSE	minimum mean squared error
OSRLS	order selective RLS
pdf	probability density function
psd	power spectral density
RBF	radial basis functions
RLP	recursive least p–norm
RLS	recursive least squares
RML	recursive maximum likelihood
RV	random variable
RWLS	recursive weighted least squares

$S\alpha S$	Symmetric α–stable

Nomenclature

$\stackrel{\mathrm{d}}{=}$	equality in distributions
$(a,\ b)$	open interval in $\mathbb{R}$; the set $\{x \in \mathbb{R} : a < x < b\}$
$[a,\ b]$	closed interval in $\mathbb{R}$; the set $\{x \in \mathbb{R} : a \leq x \leq b\}$
$(a,\ b]$	the set $\{x \in \mathbb{R} : a < x \leq b\}$
$[a,\ b)$	the set $\{x \in \mathbb{R} : a \leq x < b\}$
$\{\quad\}$	set of numbers
$[\quad]$	matrix or vector
$[\quad]^{\mathrm{T}}$	transpose of a matrix or vector
$*$	convolution operator
$(\cdot)^*$	conjugate operator
A	the binary alphabet $\{-1,\ +1\}$
$\Gamma(x)$	gamma function $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$
C_e	Euler constant $0.57721566\ldots$
$\bar{c}_i$	scalar center of the channel ($0 \leq i < 2^N$)
$\mathbf{c}_i$	vector center of the channel ($0 \leq i < 2^K$)
D	feedback order of the equaliser
$\boldsymbol{E}\{\cdot\}$	expected value
$\mathrm{erf}(x)$	the error function $\frac{2}{\sqrt{\pi}}\int_0^x \mathrm{e}^{-t^2}dt$
$\exp x$	the exponential function e^x
$f_\alpha(s)$	univariate α–stable pdf, defined in $s \in \mathbb{R}$
$f_\alpha^d(\mathbf{s})$	multivariate α–stable pdf of dimension d, defined in $\mathbf{s} \in \mathbb{R}^d$
$\mathcal{F}(\cdot)$	Fourier transform
K	length of the input vector: $K = N + M - 1$
$\ln$	the natural logarithm
L	length of the *residual* input vector: $L = K - D$
M	order of the equaliser
N	length of discrete–time channel impulse response
N_{c}	number of vector centers: $N_{\mathrm{c}} = 2^K$
N_{DFc}	number of DFE vector centers: $N_{\mathrm{DFc}} = 2^L$

N_{sc}	number of scalar centers: $N_{\text{sc}} = 2^N$
$n(k)$	noise sequence added at the output of the channel
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^m$	m–dimensional space in $\mathbb{R}$
$r(k)$	received (observation) signal sequence
$\mathbf{r}(k)$	received (observation) vector
rem	remainder after division (operator)
$x(k)$	transmitted data sequence
$\hat{x}(k)$	estimate of transmitted symbol $x(k)$
$\mathbf{x}(k)$	transmitted symbols vector
$\mathbf{x}_i$	all possible discrete states of the transmitted symbols vector $\mathbf{x}(k)$ ($0 \leq i < N_{\text{c}}$)
$\mathbf{x}_{\text{ch}}(k)$	channel input vector
$\mathbf{x}_{\text{ch}_i}$	all possible discrete states of the channel input vector $\mathbf{x}_{\text{ch}}(k)$ ($0 \leq i < N_{\text{sc}}$)
$y(k)$	noise–free output sequence of the channel
$\mathbf{y}(k)$	noise–free channel output vector
$\Phi_\alpha(\omega)$	characteristic function of stable distribution

Chapter 1
Introduction

The birth of the communications age was signalled by the invention of wire telegraphy by Morse in the 1840's. Some decades later, the transmission of voice through wires initiated the spread of wire telephony. At the beginning of the 20–th century, radio transmission was made possible with the invention of the triode tube. With the developments in the field of semiconductors, communications involving voice and video became widespread while the interest for data communications was ignited by the increasing utilisation of digital equipment such as computers, fax and mobile phones.

The theory of digital transmission was first presented by Nyquist in 1928 [2]. In 1948, Shannon [3, 4] determined the upper performance limits for transmission through a given channel with certain transmit power available. In the last decades, smart signal processing techniques have enabled the effective use of the traditional "voice grade" channels for the transmission of data. The demand for digital communications has experienced a massive growth in the last years. Moreover, communications equipment is becoming smaller, lighter and less energy consuming, while at the same time it is expected to grow more reliable, more robust, more economic in terms of spectrum use, and most important, operating at increasingly higher data rates.

Thus, an increasingly demanding and challenging telecommunications landscape is being formulated, requiring increasingly sophisticated architectures and methods for the efficient and economic utilisation of the physical transmission media.

1.1 Motivation for work

High speed data transmission over communication channels is subject to intersymbol interference (ISI) and noise. The intersymbol interference is usually the result of the restricted bandwidth allocated to the channel and/or the presence of multipath distortion in the medium through which the information is transmitted. Equalisation is the process which reconstructs the transmitted data jointly combating the ISI and the noise in the communication link.

The most simple architecture in the class of equalisers making decisions in a symbol–by–symbol basis is the linear transversal filter. The optimal solution, however, is the Bayesian approach which is known as the maximum a posteriori (MAP) symbol–by–symbol decision equaliser [5].

Although the Bayesian equaliser and its adaptive implementation has been thoroughly studied in the literature (for example see [6] and the references therein), by and large, the results are related to the assumption that the interference noise is Gaussian. However, in many physical channels, such as urban, indoor radio and underwater acoustic channels [7–9], the ambient noise is known through experimental measurements to be non-Gaussian, mainly due to the impulsive nature of man–made electromagnetic interference.

It is well known that non–Gaussian noise can cause significant performance degradation in traditional communications systems designed under the Gaussian assumption. A well known example is the matched filter for coherent reception of deterministic signals in Gaussian white noise. If the noise statistics deviate from the Gaussian model, serious degradation in performance occurs, such as increased false alarm rate or error probability [10, 11].

That means, when the performance degradation due to the ideal Gaussian assumption in a non–Gaussian environment can not be tolerated, the underlying signal processing methods must be revisited and redesigned taking into account the non–Gaussian noise statistics.

A number of models have been proposed for impulsive phenomena in communication systems, either by fitting experimental data [12–14] or based on physical grounds [8, 15]. Recently, it has been suggested [9] that the family of α–stable random variables (RV) provides an appropriate and flexible model for many impulsive phenomena, including interference in communication channels. Stable distributions share defining characteristics with the Gaussian distribution, such as the stability property and the central limit theorem.

1.2 Thesis contributions

In this thesis, the impulsive noise corrupting the communication channel is modelled as an α–stable RV. The main characteristic of non–Gaussian stable RV's is their infinite variance. However, the finite variance assumption is pivotal in most signal processing techniques for communications, rendering the stable law inappropriate as a noise modelling tool. For the

study and evaluation of systems experiencing infinite variance noise, a novel approach is proposed, which examines the signals after being passed through a finite dynamic range system. This approach enables the utilisation of the flexibility of the stable law as a model for the impulsive noise, while satisfying two very desirable features; a) it overcomes the infinite variance problem, thus most signal processing techniques can still be used, and b) it is identical to the stable law within the dynamic range of the system, providing heavier tails than the Gaussian distribution.

The optimum *maximum a posteriori probabilities* (MAP) equaliser is revisited under the light of the non–Gaussian noise model and both the feed–forward and decision feedback structures are derived. The implementation scheme for the proposed equaliser is also given, which proves to be a generalisation of the radial basis functions (RBF) network realisation of the traditional Bayesian equaliser.

For the estimation of the equaliser centers, two developments are presented in this thesis. First, a family of recursive algorithms for channel estimation in α–stable environments is presented under a unified framework, highlighting the underlying relationships between them. Furthermore, a novel clustering algorithm for the direct estimation of the centers in non–Gaussian noise environments is proposed, based on robust location estimation.

Finally, a number of practical approximation tools concerning the simulation and implementation of communication systems experiencing stable signals are derived. These approximations enable the efficient calculation of certain quantities, which are either difficult, or even impossible to compute.

1.3 Thesis layout

In this thesis, chapter 2 presents the fundamental concepts of digital communications systems. The mathematical tools for the analysis and evaluation of such systems are presented and the simplifications and assumptions taken in this thesis explained. Then, the need for equalisation of channels suffering from intersymbol interference is brought forward, while a classification of the different equalisation techniques completes this chapter.

The class of α–stable random variables is presented in chapter 3, as a modelling tool for impulsive phenomena in communications. The main characteristics and properties of stable dis-

tributions are discussed and placed into the communications context. Certain implications of the adoption of the stable model from a signal processing point of view are addressed and a new framework is proposed to overcome the problem of the infinite variance of stable signals.

In chapter 4, the optimum Bayesian feed–forward and decision feedback equalisers are derived for α–stable noise environments. The problem of evaluating communication systems in infinite variance environments is also addressed and a novel analytical framework in this direction is presented. Some preliminary experimental results are presented, showing a promising performance benefit compared with a Bayesian equaliser designed under the Gaussian assumption.

Chapter 5 discusses the problem of estimating the channel and noise characteristics in an α–stable noise environment. A family of recursive algorithms for channel identification in such environments is presented and studied. The performance of the algorithms is then experimentally assessed. For the estimation of the noise parameters, a range of techniques is studied and a number of experiments are conducted for their evaluation. After a comparative study, based on efficiency, simplicity and performance, the more appropriate algorithm for the equalisation problem is highlighted.

A complete adaptive Bayesian equaliser, consisting of a MAP detector, a channel estimator and a noise parameters estimation algorithm, is experimentally studied in chapter 6. Some useful approximations concerning the practical implementation of the equaliser are also proposed.

Finally, in chapter 7 the main contributions of the thesis are discussed, certain limitations of the work are identified, while the scope for future research is brought forward.

Chapter 2
Digital communication systems and equalisation

This thesis discusses the problem of adaptive equalisation for digital communication systems (DCS's) where the interfering noise exhibits impulsive, rather than Gaussian, characteristics. In order to establish the context and motivation of the work, it is necessary to present the fundamental concepts of digital communication systems. This chapter gives a concise description of a general DCS and brings forward the need for equalisation in order to combat the impairments of the communication medium.

First, the functional modules of a generic DCS are presented and their operation explained. A convenient but accurate mathematical model for the quantitative evaluation and analysis of communication systems is then discussed, followed by the presentation of the optimum receiver. Then a more simple equivalent discrete–time model for a communication system is adopted. The significance of intersymbol interference is highlighted and the need for equalisation is brought out. An overview of the different approaches to the equalisation problem is given in section 2.4 and finally some concluding remarks are discussed.

2.1 A generic system model

Digital communication systems are designed to transmit the information generated by a source to one or more destinations in digital form. The block diagram of a general DCS is presented in fig. 2.1 [1]. The *data source* constitutes the signal generation system that generates the information to be transmitted. Information sources may take a variety of different forms. They can be *analogue*, such as audio sources in radio broadcasting or video sources in TV broadcasting. In contrast, they can be *digital* such as binary data or ASCII characters generated by computers and storage devices (e.g. magnetic or optical disks). In modern digital communication systems all the information to be transmitted must be first converted into a sequence of digits. For

[1] Some DCSs may not have some of the blocks shown here.

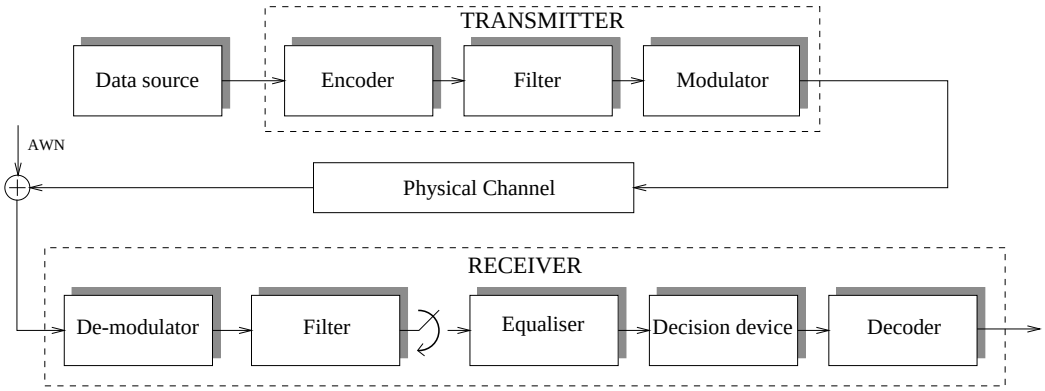


Figure 2.1: *Block diagram of a digital communication system*

non–digital sources, this is done through sampling and quantisation. Therefore, the information source can always be regarded as producing a stream of digits.

The stream of digital data, or *information sequence*, is then passed to the *encoder*. The purpose of the encoder is to introduce, in a controlled manner, some redundancy in the digital information sequence that can be used at the receiver to overcome the effects of noise and interference encountered during the transmission of the signal through the channel. This added redundancy actually serves to provide means for error detection and/or correction at the receiver end. Some of the typical coding schemes used are *Gray codes*, *block codes* (e.g., Hamming code, cyclic codes, concatenated block codes), *convolutional codes* [16] and turbo codes [17].

The information–bearing signals are usually transmitted by some type of carrier modulation. The channel over which the signal is transmitted is limited in bandwidth to an interval of frequencies centered about the carrier, as in double–sideband modulation, or adjacent to the carrier, as in single–sideband modulation. The band–limited channel has finite capacity, which for additive white Gaussian noise is given by

$$C = W \log_2 \left(1 + \frac{P_{\text{av}}}{W N_0} \right) \tag{2.1}$$

where W is the channel bandwidth, P_{av} is the average power of the transmitted signal, and N_0 is the power spectral density of the white Gaussian noise. The units of the capacity C are bits per input symbol into the channel. This formula was first derived by Shannon in [3]. Shannon [4] also showed that if the transmission rate is less than the channel capacity, then it is possible to achieve reliable communication, with as small an error probability as desired. The efficient use of this restricted bandwidth is achieved through the choice of the encoding scheme and the

design of the *transmitter filter*, also called the modulating filter.

The *modulator* on the other hand places the baseband signal over a high frequency carrier for transmission in the allocated spectrum using a modulation scheme. Some of the typical modulation schemes used in digital communication systems are amplitude shift keying (ASK), frequency shift keying (FSK), pulse amplitude modulation (PAM) and phase shift keying (PSK) modulation.

The communication *channel* is the physical medium that is used to send the high frequency carrier from the transmitter to the receiver. In wireless transmission, the channel may be the atmosphere (free space). On the other hand, telephone channels usually utilise a variety of physical media, such as wire lines, optical fibre cables, and wireless microwave links. Whatever the physical medium used for transmission of the information, the transmitted signal is corrupted in a random manner by a variety of possible mechanisms, such as additive thermal noise generated by electronic devices, man–made noise (e.g. automobile ignition noise) and atmospheric noise (e.g. electrical lightning discharges during thunderstorms). This interference is modelled as random, additive white noise (AWN) at the output of a noise–free channel. Another essential characteristic of the transmission of information through a channel is that the bandwidth allocated for the channel is often limited, resulting in the dispersion of power between neighbour symbols in the transmitted sequence. This distortion of the channel is called *intersymbol interference* (ISI).

At the receiver the signal is first *demodulated* to recover the transmitted signal in its baseband form. The demodulated signal is processed by the *receiver filter*, also called receiver demodulating filter, which, as will be shown later in this chapter, should be ideally matched to the transmitter filter and channel impulse response [2].

The output of the receiver filter is sampled at the symbol rate and the resulting discrete time signal is passed to the *equaliser*. The equaliser in the receiver removes the ISI distortion introduced due to the limited bandwidth of the channel. The *decision device* reconstructs the encoded transmitted binary sequence, based on the soft decisions made by the equaliser. Finally, the *decoder* performs the reverse operation of the encoder and reconstructs the sequence of transmitted information symbols.

[2] Normally the channel transfer function is not known to the receiver and may be non-stationary. For this reason the receiver is usually matched to the transmitter filter only.

The modulation performed at the transmitting end of the communication system to generate the bandpass signal and the demodulation performed at the receiver end to recover the digital information involve frequency translations. For mathematical convenience, it is desirable to reduce all bandpass signals and channels to equivalent lowpass signals and channels. It is well known [16] that narrowband bandpass signals and systems can be represented by equivalent lowpass signals and systems, respectively. Suppose that a real–valued signal $s(t)$ has a frequency content concentrated in a narrow band of frequencies in the vicinity of a frequency f_c. The equivalent lowpass signal is found to be

$$s_l(t) = \left[\left(\delta(t) + \frac{j}{\pi t}\right) * s(t)\right] \mathrm{e}^{-j2\pi f_c t} \tag{2.2}$$

where $*$ is the convolution operation. Accordingly, we consider a linear filter or system with impulse response $h(t)$ and frequency response $H(f)$. The equivalent lowpass system has a frequency response

$$H_l(f) = \begin{cases} H(f+f_c) & (f \geq 0) \\ 0 & (f < 0) \end{cases} \tag{2.3}$$

and an impulse response $h_l(t)$, where

$$h(t) = 2\mathrm{Re}\left[h_l(t)\mathrm{e}^{j2\pi f_c t}\right] \tag{2.4}$$

Furthermore, the output $r(t)$ of the bandpass system $H(f)$ to the bandpass input signal $s(t)$ can be simply obtained from the equivalent lowpass input signal and the equivalent lowpass impulse response of the system. Specifically

$$r_l(t) = \int_{-\infty}^{\infty} s_l(\tau) h_l(t-\tau) d\tau \tag{2.5}$$

where $r_l(t)$ is the equivalent lowpass of $r(t)$. It holds that

$$r(t) = \mathrm{Re}\left[r_l(t)\mathrm{e}^{j2\pi f_c t}\right] \tag{2.6}$$

The combination of eq. (2.5) with eq. (2.6) gives the relationship between the bandpass output signal $r(t)$ and the equivalent lowpass time functions $s_l(t)$ and $h_l(t)$. This simple relationship allows us to ignore any linear frequency translations encountered in the modulation and

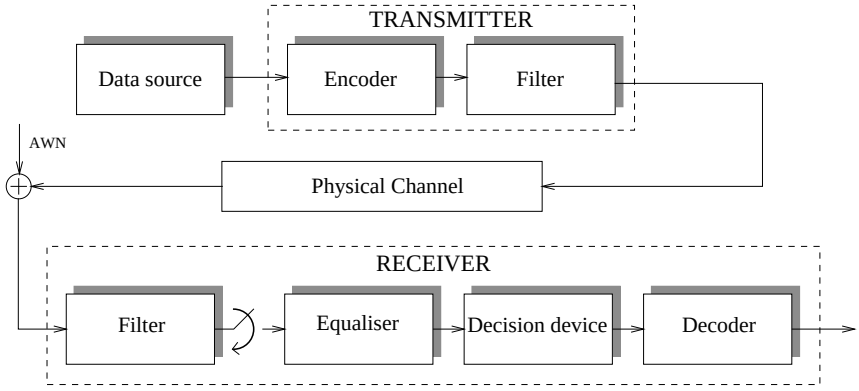


Figure 2.2: *Baseband model of digital communication system*

demodulation of the information–bearing signal. Figure 2.2 presents the equivalent baseband model of the DCS presented in fig. 2.1. Here the modulator and the demodulator have been removed.

2.2 The matched filter

As shown in [16] the equivalent lowpass transmitted signal for several digital modulation techniques has the common form

$$v(t) = \sum_{k=0}^{\infty} x_k g_{\mathrm{T}}(t - kT) \tag{2.7}$$

where $\{x_k\}$ is the encoded discrete information–bearing sequence of symbols (output of the encoder) and $g_{\mathrm{T}}(t)$ is the transmitter filter pulse, which is assumed to have a band–limited frequency response characteristic $G_{\mathrm{T}}(f)$. Here T is the symbol period, while $v(t)$ is the output of the transmitter filter in the baseband system model (fig. 2.2). Suppose that the band–limited channel is a linear filter having an equivalent lowpass frequency response characteristic $C(f)$ and an equivalent lowpass impulse response $c(t)$. Then, if the signal $v(t)$ is transmitted over the channel $C(f)$, the equivalent lowpass received signal is

$$r_{\mathrm{l}}(t) = \sum_{k=0}^{\infty} x_k h(t - kT) + n(t) \tag{2.8}$$

where

$$h(t) = \int_{-\infty}^{\infty} g_{\mathrm{T}}(\tau) c(t-\tau) d\tau \tag{2.9}$$

is the convolution of the channel impulse response with the transmitter filter pulse, and $n(t)$ represents the additive white noise. The optimum receiver consists of a receiver filter $g_{\mathrm{R}}(t)$ matched to $h(t)$ [16] followed by a sampler operating at the symbol rate $1/T$. The optimum receiver filter is

$$g_{\mathrm{R}}(t) = h^*(-t) \tag{2.10}$$

which is called the *matched filter* and its output will be

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} r_{\mathrm{l}}(\tau) g_{\mathrm{R}}(t-\tau) d\tau \\ &= \int_{-\infty}^{\infty} r_{\mathrm{l}}(\tau) h^*(\tau - t) d\tau \end{aligned} \tag{2.11}$$

Sampling $y(t)$ at the symbol rate we obtain

$$y_k \equiv y(kT) = \int_{-\infty}^{\infty} r_{\mathrm{l}}(\tau) h^*(\tau - kT) d\tau \tag{2.12}$$

If we substitute for $r_{\mathrm{l}}(\tau)$ in eq. (2.11) using eq. (2.8), we obtain

$$\begin{aligned} y_k &= \int_{-\infty}^{\infty} \left[\sum_{i=0}^{\infty} x_i h(\tau - iT) + n(\tau) \right] h^*(\tau - kT) d\tau \\ &= \sum_{i=0}^{\infty} x_i \int_{-\infty}^{\infty} h(\tau - iT) h^*(\tau - kT) d\tau + \int_{-\infty}^{\infty} n(\tau) h^*(\tau - kT) d\tau \end{aligned} \tag{2.13}$$

The second term in the right hand side of eq. (2.13) is the additive noise sequence of the output of the matched filter, that is

$$v_k \equiv v(kT) = \int_{-\infty}^{\infty} n(t) h^*(\tau - kT) d\tau \tag{2.14}$$

Furthermore, if we write the autocorrelation function of $h(t)$, we obtain

$$a_k \equiv a(kT) = \int_{-\infty}^{\infty} h(\tau) h^*(\tau - kT) d\tau \tag{2.15}$$

and it is easy to show that the integral in the first term of the right hand side of eq. (2.13) is equal to a_{k-i}. Therefore, eq. (2.13) can be rewritten as follows:

$$y_k = \sum_{i=0}^{\infty} x_i a_{k-i} + v_k \tag{2.16}$$

The output of the demodulator (matched filter) at the sampling instants kT form the sequence $\{y_k\}$ which is a set of sufficient statistics for the optimum demodulation of the transmitted signal sequence x_k [16]. However, $\{y_k\}$ is corrupted by intersymbol interference, as indicated by eq. (2.16), the extent of which depends on the autocorrelation function of $h(t)$. In any practical system, it is reasonable to assume that the ISI affects a finite number of symbols. Hence, we may assume that $a_k = 0$ for $|k| > L$, so that the ISI spans $2L + 1$ symbols.

2.3 Discrete–time model for a channel with ISI

In dealing with band–limited channels that result in ISI, it is convenient to use an equivalent discrete–time model for the continuous–time system. We have assumed that the encoder provides discrete–time symbols at a rate $1/T$ (symbols/sec) and the sampled output of the matched filter at the receiver is also a discrete–time signal, with the same symbol rate. Therefore, the cascade of the transmitting filter with impulse response $g_{\mathrm{T}}(t)$, the channel with impulse response $c(t)$, the matched filter at the receiver with impulse response $h^*(-t)$, and the sampler can be represented by an equivalent discrete–time transversal filter having $2L$ tap gain coefficients $\{a_i\}$. Its input is the sequence of information symbols from the encoder x_k and its output is the discrete–time sequence y_k given by eq. (2.16).

The major difficulty with this discrete–time model occurs in the evaluation of performance of the various equalisation or estimation techniques, due to the correlations in the noise sequence $\{v_k\}$ at the output of the matched filter. Since it is more convenient to deal with a white noise sequence, it is desirable to whiten the noise sequence by further filtering the sequence $\{y_k\}$. A discrete–time noise–whitening filter is determined as follows.

Let $A(z)$ denote the (two–sided) z transform of the sampled autocorrelation function $\{a_k\}$, i.e.,

$$A(z) = \sum_{k=-L}^{L} a_k z^{-k} \tag{2.17}$$

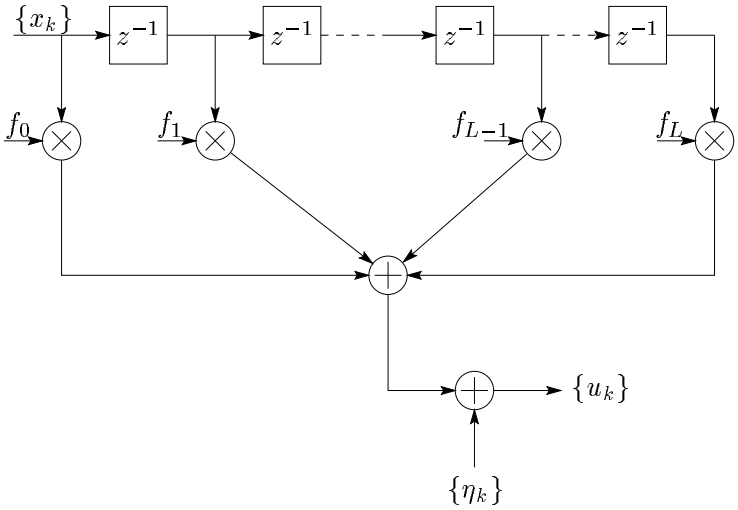


Figure 2.3: *Equivalent discrete–time model of intersymbol interference channel with white noise.*

Since $a_k = a^*_{-k}$, it follows that $A(z) = A^*(z^{-1})$ and the $2L$ roots of $A(z)$ have the symmetry that if ρ is a root, $1/\rho^*$ is also a root [16]. Hence, $A(z)$ can be factored and expressed as

$$A(z) = F(z)\ F^*(z^{-1}) \tag{2.18}$$

where $F(z)$ is a polynomial of degree L having the roots $\rho_1, \rho_2, \ldots \rho_L$ and $F^*(z^{-1})$ is a polynomial of degree L having the roots $1/\rho_1^*, 1/\rho_2^*, \ldots 1/\rho_L^*$. Then, an appropriate noise–whitening filter has a z transform $1/F^*(z^{-1})$. There are 2^L possible choices for the roots of $F^*(z^{-1})$, and each choice results in a filter characteristic that is identical in magnitude but different in phase from other choices of the roots. Choosing the unique $F^*(z^{-1})$ having minimum phase, i.e., the polynomial having all its roots inside the unit circle, then $1/F^*(z^{-1})$ is a physically realizable, stable, recursive discrete–time filter. Consequently, passage of the sequence $\{y_k\}$ through the digital filter $1/F^*(z^{-1})$ results in an output sequence $\{u_k\}$ that can be expressed as

$$u_k = \sum_{i=0}^{L} f_i x_i + \eta_k \tag{2.19}$$

where $\{\eta_k\}$ is a white noise sequence and $\{f_i\}$ is a set of $L+1$ tap coefficients of an equivalent discrete–time transversal filter having a transfer function $F(z)$.

In summary, the cascade of the transmitting filter $g_{\mathrm{T}}(t)$, the channel $c(t)$, the matched filter $h^*(-t)$, the sampler, and the discrete–time noise–whitening filter $1/F^*(z^{-1})$ can be represented as an equivalent discrete–time transversal filter having the set $\{f_i\}$ as its tap coefficients.

The additive noise sequence $\{\eta_k\}$ corrupting the output of this discrete–time transversal filter is a white noise sequence. If the underlying noise distribution is Gaussian, the sequence $\{\eta_k\}$ has zero mean. Figure 2.3 illustrates the model of the equivalent discrete–time system with white noise. This *discrete–time white noise linear filter model* of the continuous channel will be used in the remaining part of the thesis for the design and evaluation of equalisation algorithms.

2.4 Equalisation techniques

Equations 2.16 and 2.19 indicate that the output of the equivalent discrete–time system model is distorted by intersymbol interference among the data symbols. In general, all types of DCS's are affected by ISI. For example, digital transmission over analogue telephone lines experiences ISI due to the limited bandwidth of the medium. Mobile radio channels are also affected by ISI resulting from multi–path fading due to the relative motion between the transmitter and receiver [18].

The ISI may cause errors when attempting to recover the data sequence. To make things worse, the channel characteristics that cause the distortion may vary considerably in time. Therefore, it is appropriate to assume that the channel, which is modelled as a linear system, is not known during the design of the receiver. In such a case the problem is to design a corrective system which, when cascaded with the front end of the receiver produces an output that, in some sense, corrects for the distortion caused by the channel and thus yields a replica of the transmitted signal. Since the distorting system is usually unknown, it is necessary for the corrective system to identify and continuously adapt to the, often, time–varying channel. Such a system is called an *adaptive equaliser*. The equalisation problem has received great attention in the literature and different solutions to this problem may be found [19]. It is useful to present a general classification of adaptive equalisers (fig. 2.4).

In general the family of adaptive equalisers can be divided into *supervised equalisers* and *unsupervised equalisers*. For the identification of the unknown channel, it is often necessary, when possible, to periodically excite the system with a known *training* or *pilot* signal interrupting the transmission of useful information. A replica of this pilot signal is available at the receiver and the receiver compares the response of the system with its input in order to update its parameters in some manner. Such equalisers are known as supervised equalisers. However, the constraints associated with some communication systems, such as digital television or digital radio, do not

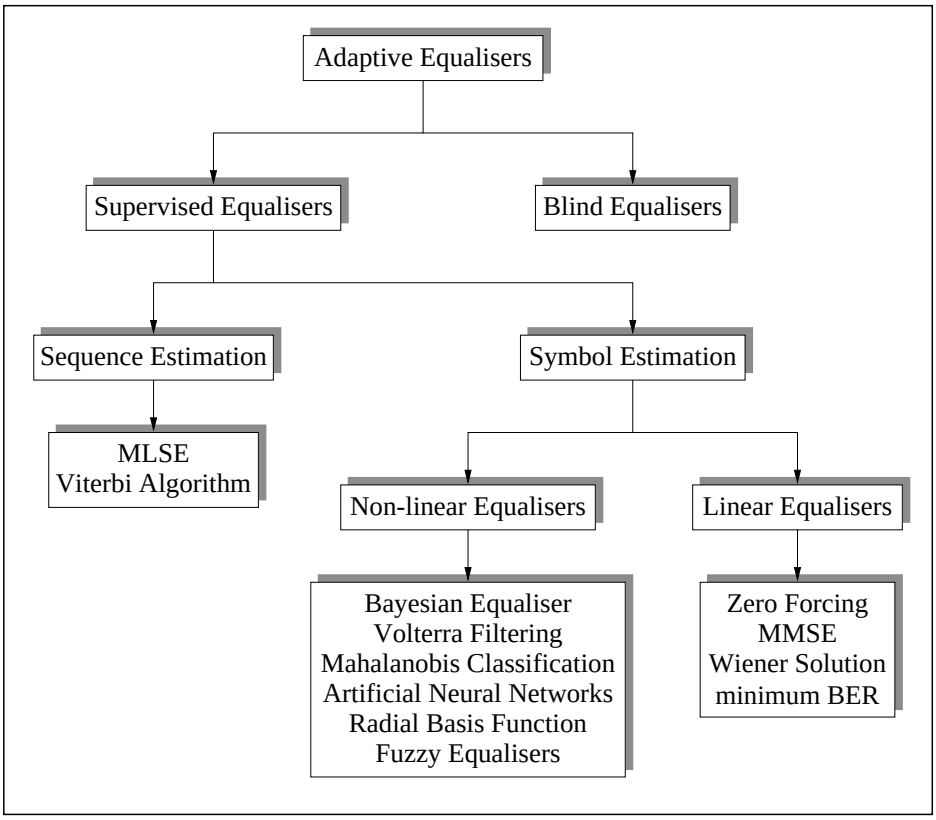


Figure 2.4: *Classification of adaptive equalisers*

provide the scope for the use of a training signal. In this situation the equaliser needs some form of unsupervised or *self recovery* method to update its parameters. These equalisers are called *blind equalisers*. After training, the equaliser is switched to *decision directed* mode, where the equaliser can update its parameters based on the actual detected data. This thesis investigates supervised equalisers only.

The process of supervised equalisation can be achieved broadly in two ways. These are *sequence estimation* and *symbol–by–symbol estimation*. The sequence estimator uses the sequence of received samples to recover the entire transmitted sequence of data symbols. The optimum sequence estimator is the maximum likelihood sequence estimator (MLSE) [20] and can be efficiently implemented based on a Viterbi trellis – the maximum likelihood Viterbi algorithm (MLVA) [21]. It is well known that the MLVA algorithm provides the best attainable equalisation performance. Since the MLSE requires that the entire data sequence has been received before the detection has been made, its theoretical performance can not be achieved in real–time systems where an arbitrary big decision lag cannot be tolerated. Hence, in practice the MLSE operates with a finite decision lag, and the bigger the lag, the more the performance

of the algorithm approaches its theoretical lower bound.

The class of symbol–by–symbol equalisers, on the other hand, detect each transmitted symbol separately. In most cases, the decision of a symbol–by–symbol equaliser can be regarded as a function of a vector containing past received samples. This *decision function* is often restricted to be linear and the resulting equaliser is referred to as a *linear equaliser*. If there are no restrictions for the decision function, the equaliser is called a *nonlinear equaliser*. The optimum decision function is in general nonlinear and is given by the maximum a posteriori probabilities (MAP) criterion derived by Bayes's theory [22]. Hence, the optimum MAP symbol–by–symbol detector (MAPSD) is also called the *Bayesian equaliser* [23]. It has been shown [24, 25] that the MAPSD provides a lower bit–error ratio for a given lag than the MLSE. At high signal to noise ratios (SNR's), their performance is virtually indistinguishable [26, 27]. On the other hand, at low SNR the MLSE is inferior to the MAPSD.

However, MLSE is popular due to a number of reasons [28]:

- In Gaussian white noise the computational complexity of the MLSE is considerably lower than the Bayesian detector – it is noteworthy that Hayes et al. [29] showed that both algorithms share the same features and that the MAPSD can be implemented in a similar manner to the familiar Viterbi algorithm trellis used for MLSE.
- The MLSE does not require knowledge of the noise variance while the MAPSD does.
- The Viterbi algorithm provides an efficient implementation for the MLSE.

Recent advances in signal processing techniques have provided a wide variety of nonlinear equalisers. These include Volterra series based equalisers [30], Mahalanobis distance equalisers [31], artificial neural networks, multi–layer perceptrons (MLP), radial basis functions (RBF) networks, fuzzy filters and fuzzy basis functions [6, 28, 32, 33]. The nonlinear equalisers, in general, treat equalisation as a pattern classification problem. This thesis addresses the problem of equalisation in impulsive noise environment from a nonlinear point of view.

A linear approach for the decision function of the symbol–by–symbol equaliser provides a computationally less complex *linear equaliser*, but at the expense of inferior performance. In order to design such linear equalisers, different optimisation criteria may be employed, such as minimum mean squared error (MMSE) or minimum amplitude distortion. The optimum, in the MMSE sense, linear equaliser is given by the Wiener equations [34], which require exact

knowledge of the channel characteristics. In practice, however, the linear equaliser is a linear filter [35] trained with an adaptive algorithm like the least mean squares (LMS) or recursive least squares (RLS). These linear equalisers treat equalisation as inverse filtering and during the process of training they optimise a certain optimisation criterion such as MMSE.

A special category of equalisers is the class of decision feedback equalisers (DFE's). The DFE uses its past decisions in order to remove part of the distorting intersymbol interference from the received signal. The transfer function of a DFE is, in general, a non–linear function of the received signal, whatever its structure, due to the feedback operation. However, the operation of the DFE can be viewed as a function computed on the samples from the received signal and past detected symbols [28]. According to the nature of this function, the DFE may be classified as either linear or non–linear. In this thesis the term nonlinear equalisers is used exclusively for those equalisers that provide a nonlinear decision function based on received samples or the received samples along with previously detected samples.

2.5 Conclusions

In this chapter the description of a generic digital communication system was given and a convenient discrete–time model was presented as a useful mathematical tool for evaluation of the subsequent work in this thesis. The distortion of the received signal due to intersymbol interference was also highlighted and the need for adaptive equalisation was underlined. Finally, we discussed the different techniques employed for the problem of equalisation, with our attention based on nonlinear equalisers.

Chapter 3
Stable random variables

In this chapter a comprehensive presentation of stable random variables as a statistical tool for modelling the impulsive noise in communication systems is given. Experimental evidence [9, 12, 36, 37] suggest that in certain communication channels the interfering noise exhibits non–Gaussian, rather than Gaussian characteristics. Section 3.1 justifies the adoption of the stable law as a model for the statistics of the noise, while section 3.2 presents the family of stable random variables and their defining characteristics. The basic properties of stable processes are given in section 3.3, while section 3.4 introduces the multivariate stable distribution. The adoption of a stable model for signals or noise has important consequences, which are analysed in section 3.5 from a signal processing point of view. The key property of stable signals is their infinite variance. This makes the use of second order moments meaningless, and weakens certain estimation tools, such as spectral analysis or least squares techniques. Moreover, a novel approach for the modelling of impulsive noise in communications is proposed, based on the finite dynamic range of the receivers. Section 3.6 introduces the *fractional lower order moments* [9] which enable signal processing in this infinite variance environment, but inevitably introduce nonlinearities.

3.1 Impulsive noise in communications

The Gaussian stochastic process has been the dominant noise model in communications and signal processing literature. In many instances the ideal Gaussian model is reasonable and can be justified by the central limit theorem [38]

Theorem 3.1 (central limit theorem) *A physical phenomenon is Gaussian if there are infinitely many independent and identically distributed (i.i.d.) contributing factors with finite variance.* □

In addition, the Gaussian assumption often leads to analytically tractable solutions for signal processing problems. However, there are physical channels in communications, such as urban

and indoor radio channels [13, 14, 39, 40] and underwater acoustic channels [41, 42], in which the ambient noise is known through experimental measurements to be decidedly non–Gaussian, or impulsive[1].

The sources of impulsive noise may be either natural such as lightning, or man made. It may include atmospheric noise in radio links, ambient acoustic noise due to ice cracking in the arctic region in underwater sonar and submarine communications. It might come from relay contacts in switches; electromagnetic devices such as elevators, printers, typewriters and copying machines; electronic apparatus such as computers, monitors and terminals; or transportation systems such as railroads and underground trains; switching transients, and accidental hits in telephone lines [9, 12, 36, 37].

Impulsive noise is more likely to exhibit sharp spikes or occasional bursts of outlying observations than one would expect from normally distributed signals. The empirical data indicate that the probability density functions (pdf's) of the impulsive noise processes exhibit a similarity to the Gaussian pdf, being bell shaped, smooth, and symmetric, but decay in the tails less rapidly than the Gaussian density function [36].

A variety of impulsive noise models can be found in literature. The most commonly used so far is the Middleton model [40], which is composed of a Rayleigh distribution for the impulse amplitude and a Poisson distribution for the occurrence of the impulses. Recently this model was enriched with new methods and results [8]. In [12] one can find a statistical description of impulsive noise on subscriber lines based on real measured data. In another quantitative approach [37] a slightly different model is proposed. Based on measurements, this model assumes a *magnitude distribution* for the observed spikes and an *occurrence probability* for the occurrence of the bursts.

Recently, however, it has been suggested [9] that the family of α–stable random variables provides useful models for a wide range of impulsive phenomena. Stable processes exhibit some very desirable properties which include:

- They can be chosen to be symmetric or asymmetric;
- They conform to the generalised central limit theorem;

[1]The term "impulsive" is used to indicate a higher probability of large interference levels compared to the Gaussian noise.

- They have heavier tails than a Gaussian process and the heaviness of their tails is controlled by a single parameter.

The theory of stable processes evolved from the investigation of the characteristic functions by Laplace and others in the 18th and 19th centuries. Although a few members of the stable family have been known for a long time, the concept of the stable distribution did not come to light until 1925, in a monograph by Lévy [43]. Since then, it has been found to provide useful models for phenomena observed in many different fields, such as economics, physics, hydrology, biology as well as electrical engineering. For detailed accounts of some of the applications of the stable law, see [44–46].

The earliest application of stable distributions was discovered in physics by the Danish astronomer Holtsmark [47]. He found that random fluctuations of gravitational fields of stars in space under certain natural assumptions are well approximated by stable processes. However, the stable distribution did not receive much serious attention until the work by Mandelbrot and his followers in economics and finance in the sixties. Because of the failure of the Gaussian assumption and least squares criterion in economic time series, Mandelbrot proposed a revolutionary approach based on stable distributions to the problem of price movement [48]. Many economical variables, such as common stock price changes, changes in speculative prices and interest rate changes have already been shown to have properties that conform closely to those of non–Gaussian stable laws [49–51].

The stable distribution has also found applications in signal processing and communications. For example, Mandelbrot and van Ness used Gaussian and stable fractional stochastic processes to describe long–range dependence arising in engineering, economics and hydrology [52]. It was also used to describe the patterns of error clustering in telephone circuits [53]. However, the most important application of the stable distribution in signal processing is in the area of impulsive noise modelling. It has recently been shown that a general class of man–made and natural impulsive noise is indeed stable under broad conditions [54]. This result has been verified by various types of experimental data. In fact, the Cauchy distribution, which is a member of the stable family, has been used in several papers such as [55] to represent severe impulsive noise. Stuck and Kleiner [56] empirically found that the noise over certain telephone lines can be best described by stable laws with characteristic exponents close to 2. They suggested that the design of receivers should take into account this noise characteristic [53].

3.2 The class of α–stable random variables

The family of stable random variables is defined as a direct generalisation of the Gaussian law and in fact includes the Gaussian density as a limiting case [2]. The symmetric stable densities have many features of the Gaussian distribution. They are smooth, unimodal, symmetric with respect to the median and bell–shaped. However, the main characteristic of a non–Gaussian stable probability density function is that its tails are heavier than those of the normal density. This is one of the main reasons why the stable law is regarded suitable for modelling signals and noise of impulsive nature. In addition, the stable distribution is very flexible as a modelling tool. It is completely determined by four parameters.

The characteristic exponent, α

The characteristic exponent controls the heaviness of the tails of the stable density and hence the impulsiveness of the respective stable process. It can take values in the interval $(0\ ,\ 2]$; a smaller value of α implies heavier tails (i.e. severe impulsiveness), while a value of α close to 2 indicates a more Gaussian type of behaviour. When $\alpha = 2$, the stable distribution is reduced to the Gaussian density.

The index of skewness, β

The index of skewness controls the symmetry of the distribution and can only take values in the interval $-1 \leq \beta \leq 1$. When $\beta = 0$, the distribution is symmetric about the centre δ. Symmetric stable distributions with characteristic exponent α are called *symmetric α–stable*, or *SαS*. If $\alpha \neq 1$, the cases $\beta > 0$ and $\beta < 0$ correspond to left skewness and right skewness, respectively. The direction of skewness is reversed if $\alpha = 1$. It is interesting to note that the effect of β decreases as $\alpha \to 2$; in fact, the Gaussian pdf is always symmetric.

The scale parameter, γ

The scale parameter, also called the *dispersion*, can be any positive number. It plays an analogous role to variance[3] and refers to the spread of the distribution. When $\alpha = 2$ the variance of the Gaussian distribution equals 2γ.

[2] In general, stable distributions are usually assumed to be non-Gaussian although the Gaussian distribution belongs to the family of stable distributions.

[3] In general (except the Gaussian case), the variance of an α–stable stochastic process is infinite.

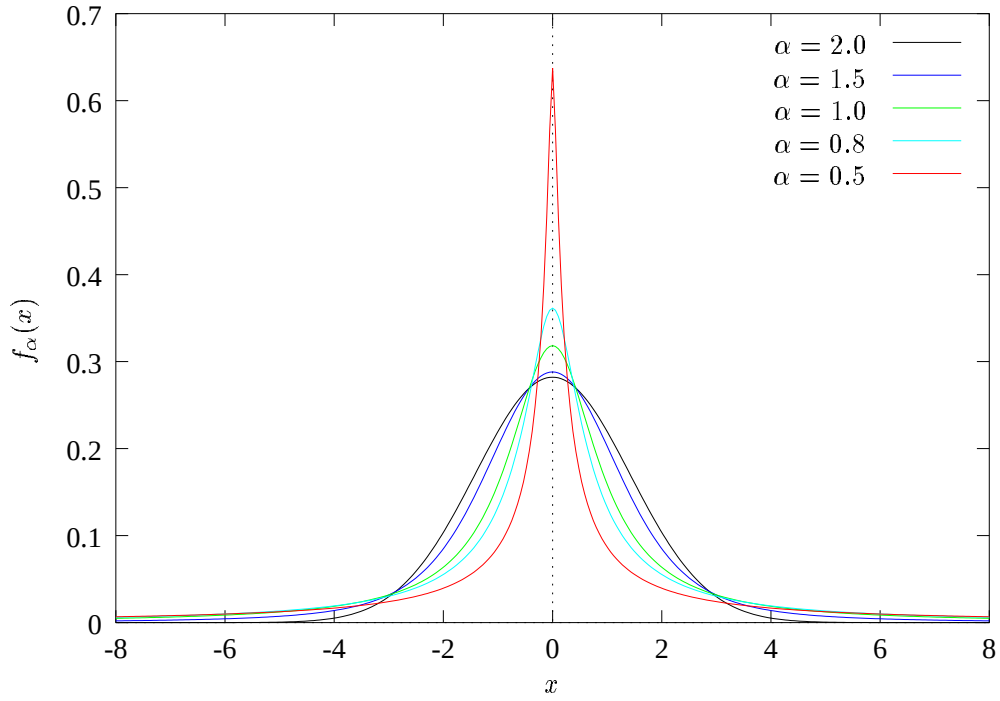


Figure 3.1: *The symmetric α–stable probability density function for five different values of the characteristic exponent α, including the Gaussian case ($\beta = 0$, $\gamma = 1$ and $\delta = 0$).*

The location parameter δ

This parameter is identical to the mean of the distribution only when $\alpha > 1$ or $\beta = 0$. If $\alpha \leq 1$ and $\beta \neq 0$ the mean of the corresponding α–stable random variable is infinite.

Figure 3.1 demonstrates the effects of α on the tails of a stable distribution. Five symmetric stable distributions are plotted, all with $\beta = 0$, $\gamma = 1$ and $\delta = 0$ but with $\alpha = 0.5, 0.8, 1.0, 1.5$ and 2.0. A detailed comparison between the normal and the stable density functions shows that non–Gaussian stable distributions depart from the corresponding Gaussian density in the following ways. For small absolute values of x, the α–stable densities are more peaked than the normal. For some intermediate range of $|x|$, the α–stable distributions have lower densities than the normal. Most importantly, unlike the Gaussian distribution, the stable densities have tails which decay less rapidly.

Figure 3.2 depicts the effects of the skewness parameter β on the symmetry of a stable distribution. Five stable densities are plotted all with $\alpha = 1.5$, $\gamma = 1$ and $\delta = 0$ but with $\beta = -1, -0.5, 0, 0.5$ and 1.

A method developed by Chambers, Mallows and Stuck [57] can be used to generate random

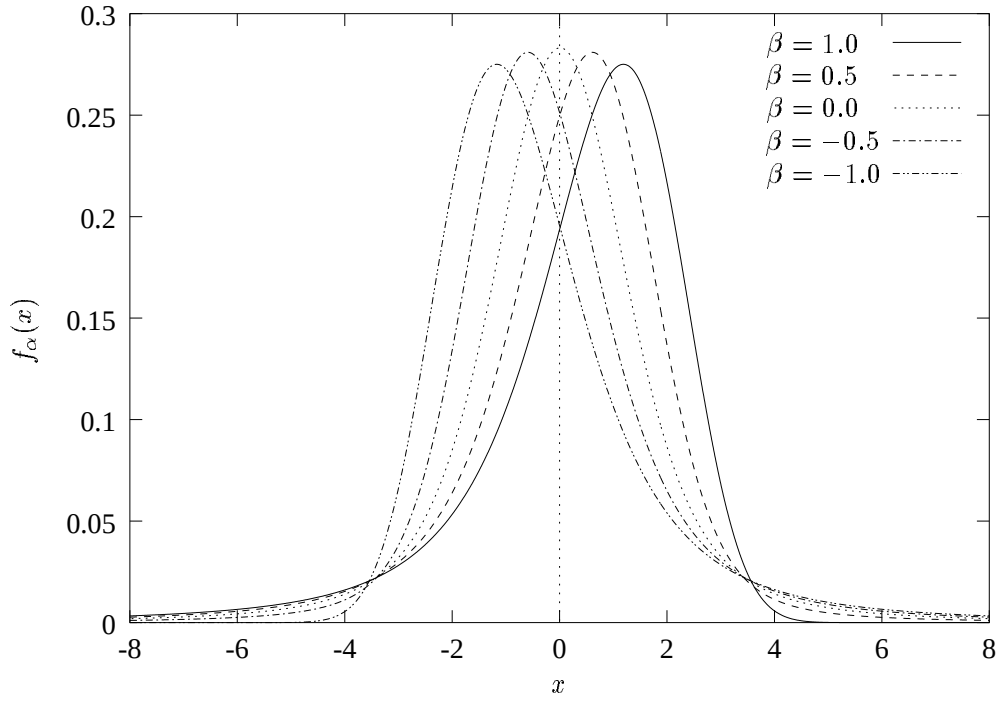


Figure 3.2: *Five α–stable distributions for a variety of values for the index of skew β ($\alpha = 1.5$, $\gamma = 1$ and $\delta = 0$).*

variables. This method was adopted in [58] and will be used throughout this thesis for the artificial generation of α–stable distributed signals. Figure 3.3 shows 1000 samples from three stable random variables with different characteristic exponent, namely $\alpha = 1.0,\ 1.5$ and 2.0. It is clear how a smaller value of α results in a higher probability of large sample levels (impulses).

The α–stable distribution $f_\alpha(s)$ of a random variable (RV) X is defined [59] by means of its characteristic function. The characteristic function $\Phi(\omega)$ of an RV X is the Fourier transform of its probability density function $f(s)$, that is

$$\Phi(\omega) = \mathcal{F}\{f(s)\} = \int_{-\infty}^{\infty} f(s)\mathrm{e}^{j\omega s} ds \tag{3.1}$$

where $\mathcal{F}\{\cdot\}$ is the operator of the Fourier transform. It is well known that the *expected value* or *mean* of the RV X is by definition the integral

$$\boldsymbol{E}\{X\} = \int_{-\infty}^{\infty} s f(s) ds \tag{3.2}$$

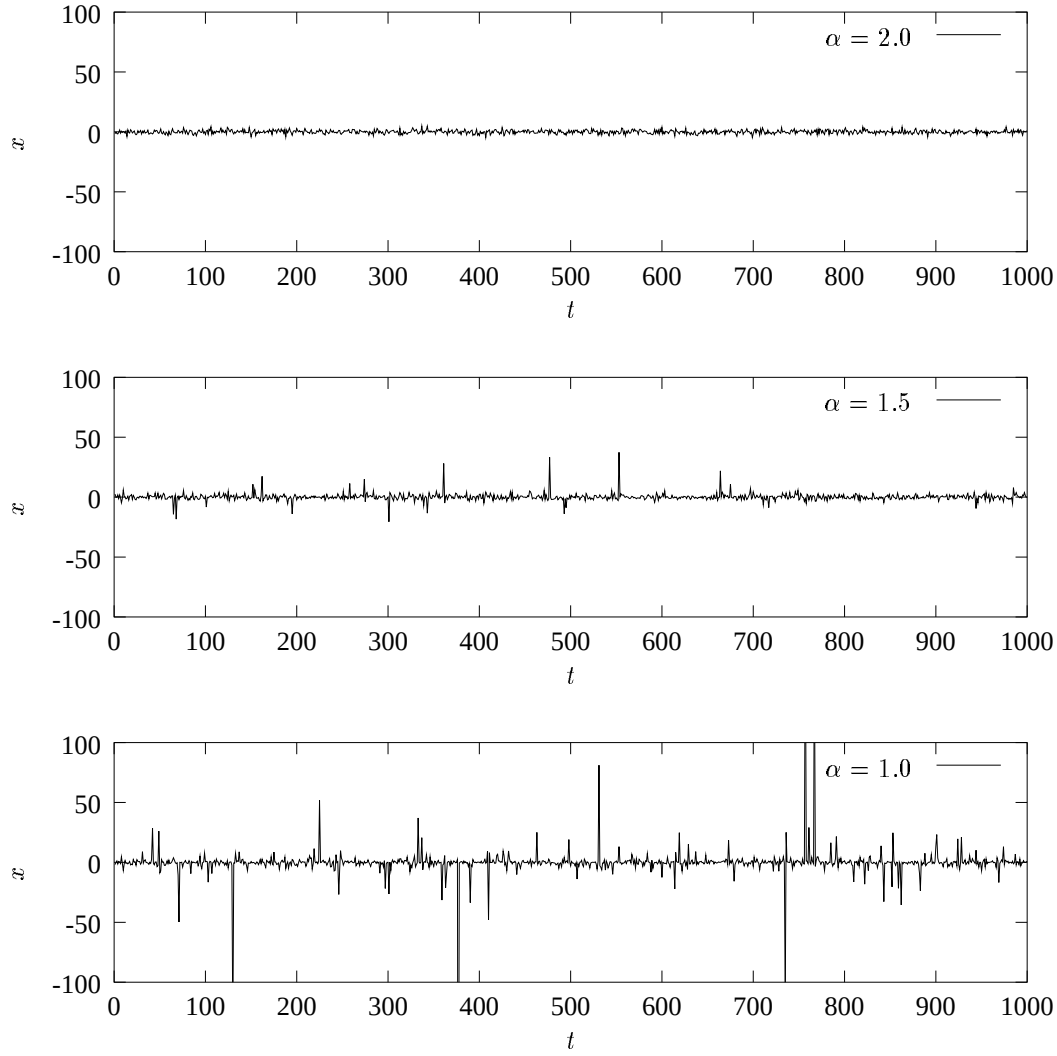


Figure 3.3: *Samples of three stable processes with* $\alpha = 1.0,\ 1.5$ *and* 2.0 *(*$\beta = 0, \gamma = 1, \delta = 0$*)*

If $g(\cdot)$ is a complex function of X

$$g(X) = g_{\mathrm{r}}(X) + jg_{\mathrm{i}}(X) \tag{3.3}$$

then, the expected value of $g(X)$ is given by [38]

$$\boldsymbol{E}\left\{g(X)\right\} = \int_{-\infty}^{\infty} g(s)f(s)ds \tag{3.4}$$

Comparing eq. (3.1) with eq. (3.4), yields an interesting property: the characteristic function

$\Phi(\omega)$ is actually the expected value of $\mathrm{e}^{j\omega X}$, i.e.,

$$\Phi(\omega) = \boldsymbol{E}\left\{\mathrm{e}^{j\omega X}\right\} \tag{3.5}$$

For a stable random variable X it can be shown [59] that its characteristic function has the following form:

$$\Phi_\alpha(\omega) = \exp\left\{j\delta\omega - \gamma|\omega|^\alpha\left[1 + j\beta\,\mathrm{sgn}(\omega)\;\rho(\omega,\alpha)\right]\right\} \tag{3.6}$$

where

$$\rho(\omega,\alpha) = \begin{cases} \tan\dfrac{\alpha\pi}{2} & : \quad \alpha \neq 1 \\ \\ \dfrac{2}{\pi}\ln|\omega| & : \quad \alpha = 1 \end{cases} \tag{3.7}$$

The α–stable distribution is the inverse Fourier transform of $\Phi_\alpha(\omega)$, and can therefore be written

$$f_\alpha(s) = \mathcal{F}^{-1}\left\{\Phi_\alpha(\omega)\right\} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi_\alpha(\omega)\;\mathrm{e}^{-j\omega s}\;d\omega \tag{3.8}$$

When the random variable is symmetric around zero ($\beta = 0$ and $\delta = 0$), eq. (3.6) is reduced to

$$\Phi_\alpha(\omega) = \exp\left(-\gamma|\omega|^\alpha\right) \tag{3.9}$$

in which case the characteristic function is real and even. That is, the density function can be simplified to

$$f_\alpha(s) = \frac{1}{\pi}\int_0^{\infty}\exp(-\gamma|\omega|^\alpha)\;\cos\omega s\;d\omega \tag{3.10}$$

A stable distribution is called *standard* if $\gamma = 1$ and $\delta = 0$. It is easy to show that if a random variable X is stable with parameters α, β, γ and δ, then

$$Y = \frac{X - \delta}{\gamma^{\frac{1}{\alpha}}} \tag{3.11}$$

is standard with characteristic exponent α and skewness parameter β, i.e.,

$$\alpha_Y = \alpha, \quad \beta_Y = \beta, \quad \gamma_Y = 1, \text{ and } \delta_Y = 0 \tag{3.12}$$

Moreover, if X is symmetric ($\beta = 0$), then the RV defined by

$$Z = cX \tag{3.13}$$

is α–stable with parameters

$$\alpha_Z = \alpha, \quad \beta_Z = 0, \quad \gamma_Z = |c|^\alpha \gamma, \text{ and } \delta_Z = c\delta \tag{3.14}$$

The probability densities of α–stable random variables exist and are continuous but, with a few exceptions, they are not known in closed form [46]. The exceptions are:

1. The Gaussian distribution ($\alpha = 2, \beta = 0$)

$$f_2(s) = \frac{1}{\sqrt{4\pi\gamma}} \exp\left(-\frac{(s-\delta)^2}{4\gamma}\right) \tag{3.15}$$

2. The Cauchy distribution ($\alpha = 1, \beta = 0$)

$$f_1(s) = \frac{\gamma}{\pi\ (\gamma^2 + (s-\delta)^2)} \tag{3.16}$$

3. The Lévy distribution ($\alpha = 0.5, \beta = 1$)

$$f_{0.5}(s) = \left(\frac{\gamma}{\sqrt{2\pi}} \frac{1}{(s-\delta)^{3/2}} \exp\left\{-\frac{\gamma^2}{2(s-\delta)}\right\}\right) \tag{3.17}$$

   which is concentrated on (δ, ∞).

However, power series expansions of stable density functions are available. The standard stable density function can be expanded into convergent series as follows [46, 60–63]. For $s > 0$,

$$f_\alpha(s) = \begin{cases} \frac{1}{\pi s} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} \Gamma(\alpha k + 1) \left(\frac{s}{r}\right)^{-\alpha k} \sin\left(\frac{k\pi}{2}(\alpha + \zeta)\right), & 0 < \alpha < 1 \\ \frac{1}{\pi s} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} \Gamma(\frac{k}{\alpha} + 1) \left(\frac{s}{r}\right)^{k} \sin\left(\frac{k\pi}{2\alpha}(\alpha + \zeta)\right) & , \quad 1 < \alpha < 2 \end{cases} \tag{3.18}$$

where

$$\begin{aligned} r &= (1+\eta^2)^{-1/(2\alpha)} \\ \eta &= \beta \tan(\pi\alpha/2) \\ \zeta &= -(2/\pi)\arctan\eta \end{aligned} \tag{3.19}$$

and $\Gamma(\cdot)$ is the usual gamma function, defined by

$$\Gamma(x) = \int_0^\infty t^{x-1}\, \mathrm{e}^{-t}\, dt \tag{3.20}$$

3.3 Basic properties of stable processes

Two of the most important properties of the stable distribution are the *stability* and the *generalised central limit theorem* [59]. They are responsible for much of the appeal of the stable distribution as a statistical model of uncertainty [9]. The stability property is actually a defining characteristic of the stable distribution and can be stated as follows:

Definition 3.1 (stability property) *A random variable X is said to have a stable distribution if for any positive numbers A and B, there is a positive number C and a real number D such that*

$$AX_1 + BX_2 \stackrel{\mathrm{d}}{=} CX + D \tag{3.21}$$

where X_1 and X_2 are independent α–stable random variables, and where the notation $X \stackrel{\mathrm{d}}{=} Y$ means that X and Y have the same distribution. □

For any stable random variable X, the characteristic exponent α is involved in eq. (3.21) in the following way

$$C^\alpha = A^\alpha + B^\alpha \tag{3.22}$$

A random variable is called *strictly* stable if eq. (3.21) holds with $D = 0$. By using the characteristic function of the stable distribution, one can easily show a more general statement: if $X_1, X_2, \ldots, X_n$ are independent and follow stable laws with the same (α, β), then all linear combinations of the form $\sum_{j=1}^n a_j X_j$ are stable with the same parameters α and β. Intuitively,

the stability property is very desirable, especially in modelling random noise and uncertain errors [9].

As a consequence of the stability property, it can be shown that stable distributions are the only possible limit distributions for sums of i.i.d random variables. This is known as the generalised central limit theorem and is formally stated as follows [64]:

Theorem 3.2 (generalised central limit theorem) *X is the limit distribution of normalised sums of the form*

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{a_n} - b_n \tag{3.23}$$

where $X_1, X_2, \ldots, X_n$ are i.i.d and $a_n \to \infty$, if and only if X is stable. □

In particular, if X_i's are i.i.d and have finite variance, then the limit distribution is Gaussian. This is of course the result of the ordinary central limit theorem.

Thus, non–Gaussian stable distributions arise as sums of random variables in the same way as the Gaussian distribution, by relaxing the finite variance constraint. If an observed signal or noise can be thought of as the sum or result of a large number of independent and identically distributed effects, then the generalised central limit theorem suggests that a stable model may be appropriate.

The main cause of the different behaviours of the Gaussian and (non–Gaussian) stable distributions is their tails. It can be shown [43, 65] that for a non–Gaussian (i.e. $\alpha < 2$) α–stable random variable X with zero location parameter and dispersion γ,

$$\lim_{\tau \to \infty} \tau^{\alpha} P(|X| > \tau) = \gamma C(\alpha) \tag{3.24}$$

where $C(\alpha)$ is a positive constant depending on α. There is a practical way to examine the behaviour of the α–stable tails, proposed as *log–tail* test by Mandelbrot [48] and also discussed in [66]. The idea is to plot $\log P(|X| > \tau)$ against $\log \tau$. Note that for zero–location symmetric densities

$$P(|X| > \tau) = 2\ P(X > \tau)$$

Since there are no closed–form expression for the stable distributions, it is only possible to

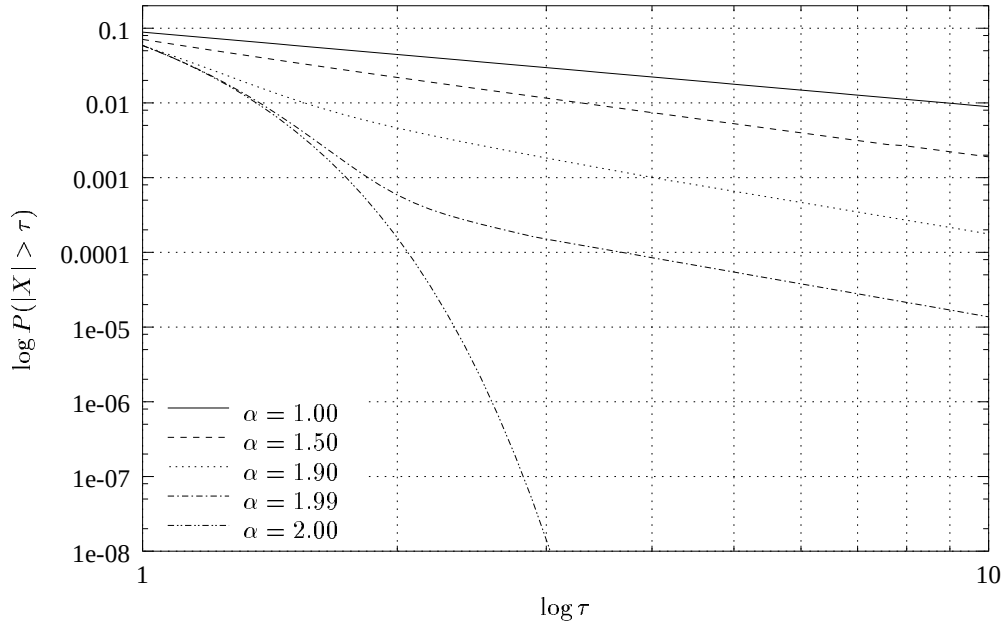


Figure 3.4: log–tail *test proposed by Mandelbrot: the probability of exceedence for a variety of values for* α *(*$\gamma = 0.14$*)*

calculate $P(X > \tau)$ for a few cases; for the zero–mean Gaussian case ($\alpha = 2, \beta = 0$ and $\delta = 0$) the probability of exceedence $P_{x>\tau}(\alpha, \gamma)$ (as a function of the noise parameters α and γ and the exceedence level τ) is

$$P_{x>\tau}(2, \gamma) = \frac{1}{2} - \frac{1}{2}\text{erf}\left(\frac{\tau}{2\sqrt{\gamma}}\right) \tag{3.25}$$

while for the zero–located Cauchy case ($\alpha = 1, \beta = 0$ and $\delta = 0$) it is

$$P_{x>\tau}(1, \gamma) = \frac{1}{2} - \frac{\arctan\left(\frac{\tau}{\gamma}\right)}{\pi} \tag{3.26}$$

In any other case, it is possible to extract this probability experimentally.

As eq. (3.24) suggests, if the α–stable distribution is non-Gaussian, the plot for $\tau \to \infty$ should be a straight line with slope $-\alpha$. Figure 3.4 depicts this test for $\alpha = 1.00,\ 1.50,\ 1.90,\ 1.99$ and 2.00. Thus stable laws have inverse power (i.e., algebraic) tails. In contrast, the Gaussian distribution has exponential tails [9]. It is noteworthy that this behaviour of the tails is obvious even when the distribution deviates slightly from Gaussian (e.g. when $\alpha = 1.99$). This proves that the tails of stable laws are much heavier than those of the Gaussian distribution. And the smaller the value of α is, the heavier the tails.

An important consequence of eq. (3.24) is the *nonexistence* of the second order moment of stable distributions, except for the limiting case $\alpha = 2$. Specifically:

Proposition 3.1 *Let X be an α–stable random variable. If $0 < \alpha < 2$, then*

$$\boldsymbol{E}\left\{|X|^p\right\} = \infty, \quad \textit{if } p \geq \alpha$$

and

$$\boldsymbol{E}\left\{|X|^p\right\} < \infty, \quad \textit{if } 0 \leq p < \alpha$$

If $\alpha = 2$, then

$$\boldsymbol{E}\left\{|X|^p\right\} < \infty, \quad \textit{for all } p \geq 0$$

□

Hence for $0 < \alpha \leq 1$, α–stable distributions have no finite first order moments; for $1 < \alpha < 2$, they have finite first order moments and all the fractional moments of order p, where $p < \alpha$; for $\alpha = 2$, all moments exist. In particular, *all non–Gaussian stable distributions have infinite variance*.

3.4 The multivariate stable distribution

The multivariate stable distribution, which is the distribution of a stable random vector, is defined by simply extending to $\mathbb{R}^d$ the definition of a stable random variable:

Definition 3.2 *A random vector $\mathbf{X} = (X_1,\ X_2,\ \ldots,\ X_d)$ is said to be a stable random vector in $\mathbb{R}^d$ if for any positive numbers A and B there is a positive number C and a vector $\mathbf{D} \in \mathbb{R}^d$ such that*

$$A\mathbf{X}^{(1)} + B\mathbf{X}^{(2)} \stackrel{\mathrm{d}}{=} C\mathbf{X} + \mathbf{D} \tag{3.27}$$

where $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are independent copies[4] *of $\mathbf{X}$.* □

[4]Consider a sequence of n independent random variables $(X_1, X_2, \ldots, X_n)$, each with the same distribution as X. We refer to these variables as independent copies of X.

As in the univariate case, C is given by

$$C^\alpha = A^\alpha + B^\alpha \tag{3.28}$$

The distribution of $\mathbf{X}$ is called a stable distribution in $\mathbb{R}^d$. Moreover, the variables X_1, X_2, ..., X_d are said to be *jointly stable*. The characteristic function of the random vector $\mathbf{X}$ can be expressed as follows

$$\Phi_\alpha^d(\boldsymbol{\omega}) = \Phi_\alpha^d(\omega_1, \omega_2, \ \ldots \ , \omega_d) = \boldsymbol{E}\left\{\exp\left(j\boldsymbol{\omega}^{\mathrm{T}}\mathbf{X}\right)\right\} = \boldsymbol{E}\left\{\exp\left(j\sum_{k=1}^{d}\omega_k X_k\right)\right\} \tag{3.29}$$

It can be shown [59] that the characteristic function $\Phi_\alpha^d(\boldsymbol{\omega})$ has the form given by the following theorem.

Theorem 3.3 *Let $0 < \alpha < 2$. Then $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ is an α–stable random vector in $\mathbb{R}^d$ if and only if there exists a finite measure Λ on the unit hyper–sphere $S_d = \{\mathbf{s} : \|s\| = 1\}$ of $\mathbb{R}^d$ and a vector $\boldsymbol{\delta}$ in $\mathbb{R}^d$ such that:*

$$\Phi_\alpha^d(\boldsymbol{\omega}) = \exp\left\{j\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\delta} - \int_{S_d}|\boldsymbol{\omega}^{\mathrm{T}}\mathbf{s}|^\alpha\left(1 - j\ \mathrm{sgn}(\boldsymbol{\omega}^{\mathrm{T}}\mathbf{s})\ \rho(\boldsymbol{\omega}, \mathbf{s}, \alpha)\right)\Lambda(d\mathbf{s})\right\} \tag{3.30}$$

□

where

$$\rho(\boldsymbol{\omega}, \mathbf{s}, \alpha) = \begin{cases} \tan\dfrac{\alpha\pi}{2} & : \quad \alpha \neq 1 \\ \dfrac{2}{\pi}\ln|\boldsymbol{\omega}^{\mathrm{T}}\mathbf{s}| & : \quad \alpha = 1 \end{cases} \tag{3.31}$$

and $\boldsymbol{\omega} \in \mathbb{R}^d$. The pair $(\Lambda, \boldsymbol{\delta})$ is unique. The measure $\Lambda(\cdot)$ is a finite Borel measure on S_d and is called the *spectral measure* of the α–stable random vector $\mathbf{X}$. In an analogy with the univariate stable characteristic function, the measure Λ controls both the scale (dispersion) and the skewness of the distribution. Except for the case $\alpha = 2$, multivariate stable distributions form a non–parametric set. They are determined by the location vector $\boldsymbol{\delta}$, the characteristic exponent $0 < \alpha \leq 2$ and the spectral measure Λ.

The following theorem is concerned with the distribution of the components of random vector $\mathbf{X}$.

Theorem 3.4 *Let* $\mathbf{X} = (X_1,\ X_2,\ \ldots\ ,\ X_d)$ *be a stable vector in* $\mathbb{R}^d$*. Any linear combination of the components of* $\mathbf{X}$ *of the type* $Y = \sum_{k=1}^{d} b_k X_k$ *is an* α*–stable random variable.* □

In the Gaussian case $\alpha = 2$, it is well known that the converse holds. For a non–Gaussian ($\alpha < 2$) random vector $\mathbf{X}$, if all linear combinations of its components are α–stable the vector $\mathbf{X}$ is, in general, not necessarily stable. However, it turns out that $\mathbf{X}$ is indeed stable if all the linear combinations are *strictly* stable, or if $\alpha \geq 1$.

Although in theory, multivariate stable distributions exist, are absolutely continuous and have continuously differentiable densities, no closed–form expressions exist for the density functions. On the other hand, the multivariate characteristic function is quite intractable, because it involves an integration over the unit hyper–sphere in $\mathbb{R}^d$. However, there is an important theorem [59] which states that

Theorem 3.5 *An* α*–stable random vector* $\mathbf{X} = (X_1,\ X_2,\ \ldots\ ,\ X_d)$ *has independent components if and only if its spectral measure* Λ *is discrete and concentrated on the intersection of the axes with the unit hyper–sphere* S_d*.* □

Suppose, for example, that $d = 2$ and consider two independent stable random variables X_1 and X_2. The measure Λ is then concentrated on the points $\mathbf{s}_1 = (1, 0)$, $\mathbf{s}_2 = (-1, 0)$, $\mathbf{s}_3 = (0, 1)$ and $\mathbf{s}_4 = (0, -1)$, i.e.,

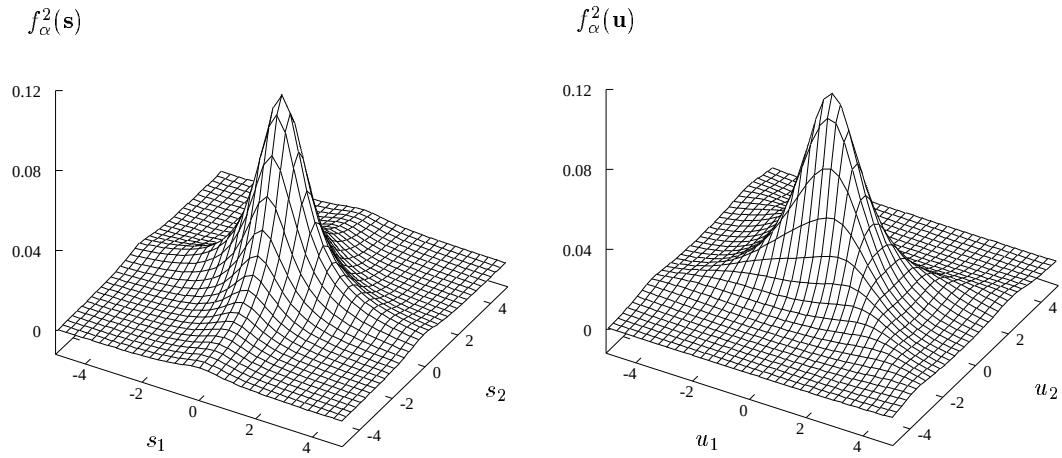
$$\Lambda = a_1\delta(\mathbf{s}_1) + a_2\delta(\mathbf{s}_2) + a_3\delta(\mathbf{s}_3) + a_4\delta(\mathbf{s}_4) \tag{3.32}$$

In the general d–dimensional case

$$\Lambda = \sum_{i=1}^{2d} a_i\delta(\mathbf{s}_i) \tag{3.33}$$

where a_i ($i = 1,\ 2,\ \ldots\ ,\ 2d$) are non–negative numbers and $\delta(\mathbf{s})$ assigns unit mass to the point $\mathbf{s}$. The points $\mathbf{s}_i$ are the $2d$ intersections of the unit hyper–sphere with the d axes. The integration over the unit hyper–sphere now becomes a summation over the points $\mathbf{s}_i$ and the characteristic function (eq. (3.30)) becomes

$$\Phi_\alpha^d(\boldsymbol{\omega}) = \exp\left\{ j(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\delta}) - \sum_{i=1}^{2d} |\boldsymbol{\omega}^{\mathrm{T}}\mathbf{s}_i|^\alpha (1 - j\,\mathrm{sgn}(\boldsymbol{\omega}^{\mathrm{T}}\mathbf{s}_i))\rho(\boldsymbol{\omega}, \mathbf{s}_i, \alpha)a_i \right\}$$



(a) Vector $\mathbf{Y}$ with independent components

(b) Vector $\mathbf{Z}$ with statistically dependent components

Figure 3.5: *The* 2*–dimensional multivariate distribution of stable random vectors.*

Equating the characteristic function to

$$\begin{aligned}
\boldsymbol{E}\left\{\exp\left(j\boldsymbol{\omega}^{\mathrm{T}}\mathbf{X}\right)\right\} &= \boldsymbol{E}\left\{\exp\left(j\sum_{k=1}^{d}\omega_k X_k\right)\right\} = \boldsymbol{E}\left\{\prod_{k=1}^{d}\exp\left(j\omega_k X_k\right)\right\} \\
&= \prod_{k=1}^{d}\boldsymbol{E}\left\{\exp\left(j\omega_k X_k\right)\right\} = \prod_{k=1}^{d}\Phi_\alpha(\omega_k) \qquad (3.34)\\
&= \prod_{k=1}^{d}\exp\left\{j\delta_k\omega_k - \gamma_k|\omega_k|^\alpha\left(1 + j\beta_k\mathrm{sgn}(\omega_k)\rho(\omega_k,\alpha)\right)\right\} \\
&= \exp\left\{\sum_{k=1}^{d}\left[j\delta_k\omega_k - \gamma_k|\omega_k|^\alpha\left(1 + j\beta_k\mathrm{sgn}(\omega_k)\rho(\omega_k,\alpha)\right)\right]\right\}
\end{aligned}$$

yields

$$\begin{aligned}
a_{2i-1} &= \gamma_i\frac{1+\beta_i}{2} \\
a_{2i} &= \gamma_i\frac{1-\beta_i}{2}, \quad i = 1,\,2,\,\ldots,\,d \qquad (3.35)\\
\boldsymbol{\delta}_i &= \delta_i
\end{aligned}$$

where $(\alpha, \beta_k, \gamma_k, \delta_k)$ are the parameters for the stable distribution of the component X_k. Here, $\boldsymbol{\delta}_i$ is the i–th element of the location vector $\boldsymbol{\delta}$, while δ_i is the location of the i–th component of the random vector $\mathbf{X}$. This proves the following theorem:

Theorem 3.6 *Let* $\mathbf{X} = (X_1,\ X_2,\ \ldots\ ,\ X_d)$ *be a stable vector in* $\mathbb{R}^d$ *with independent components. Then the components of the location vector* $\boldsymbol{\delta}$ *are equal to the location parameters* δ_k *(*$k = 1,\ 2,\ \ldots\ ,\ d$*) of the components* $X_1, X_2, \ldots, X_d$. □

To summarise, when the components of a stable random vector are independent, the characteristic function can be factorised as a product of the characteristic functions of the components (eq. (3.34)). Therefore, the stable distribution in $\mathbb{R}^d$ can be expressed as follows

$$\begin{aligned} f_\alpha^d(\mathbf{s}) &= \mathcal{F}^{-1}\left\{\Phi_\alpha^d(\boldsymbol{\omega})\right\} = \mathcal{F}^{-1}\left\{\prod_{k=1}^{d}\Phi_\alpha(\omega_k)\right\} \\ &= \left(\frac{1}{2\pi}\right)^d \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \mathrm{e}^{-j\boldsymbol{\omega}^{\mathrm{T}}\mathbf{s}} \prod_{k=1}^{d}\Phi_\alpha(\omega_k)\,d\omega_1 d\omega_2\cdots d\omega_d \\ &= \prod_{k=1}^{d}\frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi_\alpha(\omega_k)\mathrm{e}^{-j\omega_k s_k}\,d\omega_k \\ &= \prod_{k=1}^{d} f_\alpha(s_k;\beta_k,\gamma_k,\delta_k) \end{aligned} \tag{3.36}$$

where $f_\alpha(s_k;\beta_k,\gamma_k,\delta_k)$ is the α–stable distribution of the component X_k. Thus, we obtain an important theorem:

Theorem 3.7 *If* $\mathbf{X}$ *is a stable vector in* $\mathbb{R}^d$ *with independent components* $X_1, X_2, \ldots, X_d$*, then the stable distribution of* $\mathbf{X}$ *in* $\mathbb{R}^d$ *is the product of the stable distributions of the components* $f_\alpha(s_k;\beta_k,\gamma_k,\delta_k)$. □

This theorem is very useful for signal processing in α–stable environments, as will be seen in chapter 4, because it provides the means for evaluating the distribution of the vector of received samples.

As an example, consider the stable random vector $\mathbf{Y} = (Y_1,\ Y_2)$ in $\mathbb{R}^2$. Its components are i.i.d. α–stable random variables with distribution $f_{Y_k}(s) = f_\alpha(s)$ ($k = 1,\ 2\ ,\ s \in \mathbb{R}$) with $\alpha = 1, \beta = 0, \gamma = 1$ and $\delta = 0$. The distribution of vector $\mathbf{Y}$ is shown in fig. 3.5(a) and as eq. (3.36) suggests, it is given by

$$f_{\mathbf{Y}}(\mathbf{s}) = f_\alpha(s_1)\ f_\alpha(s_2) \quad ,\quad \forall\ \mathbf{s} = [s_1,\ s_2] \in \mathbb{R}^2 \tag{3.37}$$

Suppose now that the vector $\mathbf{Z} = (Z_1,\ Z_2)$ is formed by pre–multiplying vector $\mathbf{Y}$ with the

matrix

$$\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 0.5 & 0.75 \end{bmatrix} \tag{3.38}$$

i.e., $\mathbf{Z} = \mathbf{AY}$. Thus, the components of $\mathbf{Z}$ are linear combinations of Y_1 and Y_2, and specifically

$$Z_1 = \mathbf{A}_{1,1}Y_1 + \mathbf{A}_{1,2}Y_2$$
$$Z_2 = \mathbf{A}_{2,1}Y_1 + \mathbf{A}_{2,2}Y_2$$

According to theorem 3.4 (page 30) the distribution of both Z_1 and Z_2 is α–stable with the same characteristic exponent as $\mathbf{Y}$. Since Y_1 and Y_2 are independent, eq. (3.14) suggests that the dispersion of Z_1 and Z_2 will be

$$\begin{aligned} \gamma_{Z_1} &= |\mathbf{A}_{1,1}|^\alpha \gamma_{Y_1} + |\mathbf{A}_{1,2}|^\alpha \gamma_{Y_2} = 2 \\ \gamma_{Z_2} &= |\mathbf{A}_{2,1}|^\alpha \gamma_{Y_1} + |\mathbf{A}_{2,2}|^\alpha \gamma_{Y_2} = 1.25 \end{aligned}$$

The multivariate distribution of $\mathbf{Z}$ can theoretically be computed taking the inverse Fourier transform of eq. (3.30). But, although theorem 3.3 (page 30) ensures the existence of the Borel measure Λ, it does not provide its form. So, the computation of the multivariate distribution through its characteristic function becomes quite intractable. However, when the dependency between the components of an α–stable random vector is linear (as is the case with vector $\mathbf{Z}$) and the operation is reversible (i.e., matrix $\mathbf{A}$ is invertible), we will show that the α–stable multivariate distribution of the vector can be considerably simplified.

The distribution of $\mathbf{Y}$ can be expressed as $f_{\mathbf{Y}}(\mathbf{s})$, defined in the space $\mathbf{s} \in \mathbb{R}^d$. Then, since $\mathbf{Z} = \mathbf{AY}$, the distribution of $\mathbf{Z}$ defined in the space $[\mathbf{As} : \mathbf{s} \in \mathbb{R}^d]$ is identical to $f_{\mathbf{Y}}(\mathbf{s})$, i.e.,

$$f_{\mathbf{Z}}(\mathbf{As}) = f_{\mathbf{Y}}(\mathbf{s}) \tag{3.39}$$

Setting now $\mathbf{u} = \mathbf{As}$ and assuming $\mathbf{A}$ is invertible, we obtain

$$f_{\mathbf{Z}}(\mathbf{u}) = f_{\mathbf{Y}}(\mathbf{A}^{-1}\mathbf{u}) \quad , \quad \forall \mathbf{u} \in \mathbb{R}^d \tag{3.40}$$

That is, the distribution of vector $\mathbf{Z}$ in space $\mathbf{u} \in \mathbb{R}^d$ is identical with the distribution of $\mathbf{Y}$ in

the transformed space $[\bar{\mathbf{u}} = \mathbf{A}^{-1}\mathbf{u} : \mathbf{u} \in \mathbb{R}^d]$. Then, the multivariate distribution of $\mathbf{Z}$ can be written as follows

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{u}) &= \prod_{k=1}^{d} f_\alpha(\bar{u}_k) \\ \bar{\mathbf{u}} &= \mathbf{A}^{-1}\mathbf{u} \end{aligned} \tag{3.41}$$

To elaborate the above example, the inverse of matrix $\mathbf{A}$ is

$$\mathbf{A}^{-1} = \begin{bmatrix} 0.6 & 0.8 \\ -0.4 & 0.8 \end{bmatrix} \tag{3.42}$$

Thus, the transformed space $\bar{\mathbf{u}}$ will be

$$\bar{\mathbf{u}} = \begin{bmatrix} \bar{u}_1 \\ \bar{u}_2 \end{bmatrix} = \begin{bmatrix} 0.6 & 0.8 \\ -0.4 & 0.8 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0.6u_1 + 0.8u_2 \\ -0.4u_1 + 0.8u_2 \end{bmatrix} \tag{3.43}$$

and the distribution of vector $\mathbf{Z}$ can be written as

$$f_{\mathbf{Z}}(\mathbf{u}) = f_\alpha(0.6u_1 + 0.8u_2)\; f_\alpha(-0.4u_1 + 0.8u_2) \quad , \quad \forall\, \mathbf{u} = [u_1,\, u_2] \in \mathbb{R}^2 \tag{3.44}$$

This distribution is depicted in fig. 3.5(b).

3.5 Signal processing and stable distributions

From the signal processing point of view, the adoption of a stable model for signals or noise has important consequences. It has been shown that, for a non–Gaussian stable distribution with characteristic exponent α, only moments of order less than α are finite. In particular, the variance (i.e. the second order moment) of a stable distribution with $\alpha < 2$ does not exist, making the use of variance as a measure of dispersion meaningless. Similarly, many standard signal processing tools (e.g. spectral analysis and all least squares techniques) which are based on the assumption of finite variance will be considerably weakened and may in fact give misleading results.

Specifically, the additive white noise (AWN) in a digital communication system (DCS) (see fig. 2.1, page 6) is often assumed to be Gaussian. This assumption is pivotal because the finite

variance of the Gaussian noise enables the spectral characterisation of the system, which is based on second order moments. So, it is possible to reduce the passband model of the DCS to its baseband equivalent. Moreover, the design of the optimum receiver filter is based on the second order statistics of the received signal, as well as the whitening filter after the sampler. Thus, assuming that the variance of the AWN is finite it is possible to reduce the DCS into a sample–rate, discrete–time equivalent system.

Adopting the stable law as a characterisation model for the AWN at the output of the physical channel, makes the above reduction mathematically invalid, because of the infinite variance of stable RV's. Therefore, if we assume that the AWN in fig. 2.1 is α–stable, the noise samples at the output of the discrete–time equivalent system designed as described in chapter 2 will not be necessarily an independent α–stable sequence.

At this point, it should be noted that α–stable random variables are mathematical entities which provide models for heavy–tailed distributions, but represent signals that can not be physically realised or studied. The most obvious analogy is the ideal white noise, with constant power density throughout the spectrum. This is not a physically realisable signal with infinite power, and we can only study its properties when passed through a band limited system. Accordingly, we maintain that to study infinite variance signals, someone has to examine them after being passed through a *finite dynamic range* system. Besides, in practice all communication systems have a finite dynamic range, just as they have finite bandwidth.

In this thesis we assume that the transfer function of the front end of the receiver is that of an ideal saturation device, i.e.,

$$g(x,G) = \begin{cases} x & : \quad |x| \leq G \\ \mathrm{sgn}(x)\ G & : \quad \text{elsewhere} \end{cases} \tag{3.45}$$

where G is the saturation level. The operation of the ideal saturation device is to restrict the received signal within the range $[-G,\ G]$. Hence, the power contained in the tails of the received signal distribution (i.e., over G or below $-G$) is concentrated at the saturation points in the pdf as Dirac pulses. The resulting *truncated* signal possesses finite variance, and its probability density function, within the range $(-G,\ G)$, is identical to that of the unlimited signal. The relative amplitude of the Dirac pulses on the points $-G$ and G depends only a) on the shape of the distribution within the dynamic range, and b) on the saturation level G.

That is, the shape of the tails in the distribution of the received signal (before the saturation device) has no effect on the distribution of the signal at the output of the saturation device. This argument, enables us to utilise the family of stable distributions in order to model the distribution of the received signal within the dynamic range of a receiver. Thus, any distribution which is identical to the stable distribution within the dynamic range of the receiver would be adequate to describe the pdf of the receiving signal. In this family of distributions there are certain members with finite variance (the most obvious example is the *truncated* α–stable distribution within the range $[-H,\ H]$, for any $H > 2G$).

In this thesis, we assume that the additive noise at the output of the physical channel (fig. 2.1 on page 6) follows *any* distribution with finite variance from the aforementioned family of distributions, so that the mathematical reduction of the DCS to a discrete–time baseband equivalent model is meaningful and valid. By so doing, the discrete–time sequence at the output of the sampler exhibits stable characteristics within the finite dynamic range of the receiver. Furthermore, for mathematical convenience it is useful to consider the noise content at the output of the whitening filter to be statistically independent. The mathematical description of the distribution of the discrete–time received sequence will be given in section 4.5.

Adopting a non–Gaussian noise model, also implies a different approach to the design of the receiver. The efficiency of the least squares criterion and techniques, for example, is questionable under the stable assumption. The stable distribution is best used to model signals and noise that exhibit impulsive nature. This type of signals tends to produce outliers. Although the least squares criterion is adequate under the Gaussian assumption and often leads to analytically tractable results, it is no longer appropriate for an impulsive non–Gaussian environment, largely due to its non–robustness against outliers [67, 68]. It has been demonstrated many times in the literature that least squares estimates can deteriorate dramatically when only a small proportion of extreme observations is present in the data [69].

The weakness of the variance, and in general second order moments, in stable environments, implies that other measures of *variability* are needed for stable random variables. As shown in [9], the *dispersion* of a stable random variable plays an analogous role of the variance. For example, the larger the dispersion of a stable distribution, the more it spreads around its median. Hence the *minimum dispersion* criterion becomes a natural and mathematically meaningful choice as a measure of optimality in stable signal processing. By minimising the error dispersion, the average magnitude of the estimation error is minimised as well. Furthermore, it has

been shown that minimising the dispersion is also equivalent to minimising the probability of large estimation errors [70]. The minimum dispersion criterion is thus well justified under the stable assumption. It is a generalised version of the minimum mean squared error criterion (they are the same in the Gaussian case) and reasonably simple to calculate.

Minimising the dispersion is also equivalent to minimising the fractional lower order moments of estimation errors which measure the L_p distance between an estimate and its true value, for $p < \alpha \leq 2$. This result is not surprising since the L_p norm with $p < 2$ is well known for being robust against outliers such as those that may be described by the stable law [71–73]. It is also known that all lower order moments of a stable RV are equivalent, i.e. any two of the lower order moments differ by a fixed constant which is independent of the random variable itself. A common choice is the L_1 norm, which is often very convenient.

Stable signal processing based on fractional lower order moments, however, inevitably introduces nonlinearity to even linear problems. The basic reason for the nonlinearity is that we have to solve linear estimation problems in Banach or metric spaces instead of Hilbert spaces. It is well known that, while the linear space generated by a Gaussian process is a Hilbert space, the linear space of a stable process is a Banach space when $1 \leq \alpha < 2$ and only a metric space when $0 < \alpha < 1$ [74]. Banach or metric spaces do not have as nice properties and structures as Hilbert spaces for linear estimation problems.

For example, while any finite number of Gaussian random variables can be expressed as linear combinations of *independent* Gaussian random variables, it is shown in [75] that it is impossible to represent even two stable random variables of the same characteristic exponent as linear combinations of finitely many independent stable random variables. Despite the aforementioned difficulties, significant progress has been made in developing linear estimation theory for stable processes [9].

3.6 Fractional lower order moments

Although the second order moment of an α–stable random variable with $0 < \alpha < 2$ does not exist, all moments of order less than α do exist and are called *fractional lower order moments* or FLOM's [9]. The FLOM's of a $S\alpha S$ random variable can be easily found from its dispersion and characteristic exponent, from an important result which was first proved by Zolotarev [46]:

Theorem 3.8 *Let X be a* SαS *random variable with zero location parameter and dispersion γ. Then,*

$$\boldsymbol{E}\left\{|X|^p\right\} = C(p,\alpha)\gamma^{\frac{p}{\alpha}} \quad \textit{for } 0 < p < \alpha \tag{3.46}$$

where

$$C(p,\alpha) = \frac{2^{p+1}\Gamma(\frac{p+1}{2})\Gamma(-p/\alpha)}{\alpha\sqrt{\pi}\Gamma(-p/2)} \tag{3.47}$$

depends only on α and p, not on X. □

Let X be a *SαS* random variable with dispersion $\gamma > 0$ and location parameter $\delta = 0$. The norm of X is defined as

$$\| X \|_\alpha = \begin{cases} \gamma^{\frac{1}{\alpha}} & , \quad \text{for } 1 \leq \alpha \leq 2 \\ \gamma & , \quad \text{for } 0 < \alpha < 1 \end{cases} \tag{3.48}$$

Thus the norm $\| X \|_\alpha$ is basically a "scaled" version of the dispersion. If X, Y are jointly *SαS*, the distance between X and Y is defined as

$$d_\alpha(X,Y) = \| X - Y \|_\alpha \tag{3.49}$$

Combining eq. (3.46) and eq. (3.48), one can easily see that [9]

$$d_\alpha(X,Y) = \begin{cases} \left(\dfrac{\boldsymbol{E}\left\{|X-Y|^p\right\}}{C(p,\alpha)}\right)^{1/p} & , \quad \text{for } 0 < p < \alpha, \quad 1 \leq \alpha \leq 2 \\ \left(\dfrac{\boldsymbol{E}\left\{|X-Y|^p\right\}}{C(p,\alpha)}\right)^{\alpha/p} & , \quad \text{for } 0 < p < \alpha, \quad 0 < \alpha < 1 \end{cases}$$

Hence, the distance d_α between two *SαS* random variables measures the pth order moment of the difference of these two random variables for any $0 < p < \alpha$. In the case of $\alpha = 2$, this distance is half of the variance of the difference. In addition, all lower order moments of a *SαS* random variable are equivalent, i.e., the pth and qth order moments differ by a constant factor independent of the *SαS* random variable for all $0 < p, q < \alpha$.

It should also be mentioned that convergence in distance d_α is equivalent to convergence in probability [74]: a sequence of *SαS* random variables X_n converges to a *SαS* random variable Y in d_α if and only if it converges to Y in probability.

3.7 Conclusions

This chapter supports the use of the stable law as a natural generalisation of the central limit theorem in order to model random signals exhibiting impulsive behaviour. The stable law defines a family of distributions with tails that are heavier than those of the normal density, and the shape of these tails is controlled by a single parameter. Unfortunately, no closed–form expressions exist for the stable distributions, but numerical approximation can be used. The most important property of stable random variables is that they have infinite variance. On the other hand, the multivariate stable distribution is, to a great extent, mathematically intractable. However, it can be easily shown that, for vectors with independent components, the multivariate distribution can be considerably simplified.

Adopting the stable law as a characterisation model for the additive noise makes the traditional reduction of a digital communications system into a discrete–time baseband equivalent mathematically invalid. For the study of systems experiencing infinite variance noise, we propose a novel approach, which examines the signals after being passed through a finite dynamic range system. This approach enables us to choose a suitable model for the impulsive noise, which, at the same time satisfies two very desirable features; a) it has finite variance, making the discrete–time baseband equivalent of a DCS mathematical meaningful, and b) it is identical to the stable law within the dynamic range of the receiver, thus providing heavier tails than the Gaussian distribution.

The non–Gaussian noise model, also, has significant implications in signal processing algorithms based on second order moments and quadratic optimisation criteria. The fractional lower order moments provide a useful and powerful framework for optimisation in infinite power environments. However, their use implies that linear estimation problems can not be solved in Hilbert spaces which provides the scope for research towards suboptimal solutions.

Chapter 4
Bayesian equalisation

It is well known that the optimum symbol–by–symbol equaliser is the one designed under the maximum *a posteriori* probability (MAP) criterion which is derived from Bayes' theory [22]. This equaliser is also known as the Bayesian equaliser and has recently received attention in the communications literature [6, 26, 27]. However, by and large, most of the results are related to the assumption that the interfering noise exhibits Gaussian characteristics. In this chapter we derive the Bayesian equaliser for environments where the noise follows the stable law, which was presented in Chapter 3.

First, the optimum (in the MAP criterion sense) symbol–by–symbol feed–forward equaliser is derived for α–stable noise environments. Next, this equaliser is extended to a more general decision feedback Bayesian equaliser. The latter uses its past detected symbols to remove some of the intersymbol interference from the receiving signal in order to enhance its performance. In a stable noise environment, which in general implies that the noise has infinite power, the traditional performance measures based on the signal to noise ratio are meaningless. For the quantitative evaluation of equalisers in α–stable environments a novel framework is subsequently proposed. With the use of this framework a number of experiments are performed assuming that the equaliser has perfect knowledge of the channel and noise characteristics, and their results are presented in section 4.6. Finally, some concluding remarks summarise this chapter.

4.1 System model

The discrete–time model of the communication channel adopted in this thesis was depicted in fig. 2.3 (page 12). Specifically, we assume that the information source produces a digital data sequence $\{x(k)\}$, which is independently identically distributed and drawn with equal probability from a binary alphabet $A = \{-1, +1\}$. This sequence is passed through a noiseless linear discrete–time dispersive channel with finite impulse response (FIR) which spans over N

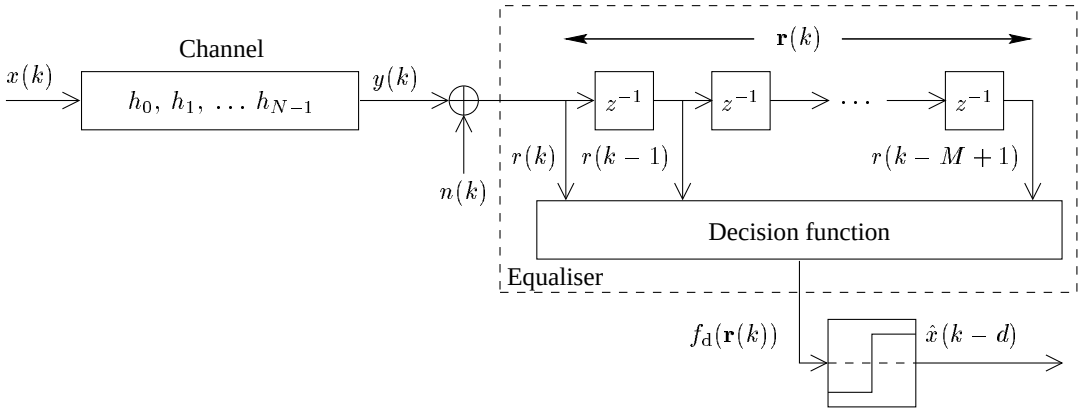


Figure 4.1: *System model for FIR channel and finite memory equaliser.*

symbols. Suppose that the z transform of the channel response can be written as

$$H(z) = \sum_{i=0}^{N-1} h_i z^{-i} \tag{4.1}$$

where

$$\mathbf{h} = [h_0\ h_1\ \ldots\ h_{N-1}]^{\mathrm{T}} \tag{4.2}$$

is the vector of channel tap–weights. If

$$\mathbf{x}_{\mathrm{ch}}(k) = [x(k)\ x(k-1)\ \ldots\ x(k-N+1)]^{\mathrm{T}} \tag{4.3}$$

is the channel input vector, then the output $y(k)$ of the channel will be the convolution of the channel tap–weight vector $\mathbf{h}$ with the channel input vector $\mathbf{x}_{\mathrm{ch}}(k)$, i.e.,

$$y(k) = \mathbf{h}^{\mathrm{T}}\mathbf{x}_{\mathrm{ch}}(k) \tag{4.4}$$

In this thesis we assume that the random additive white noise sequence $\{n(k)\}$ at the output of the channel exhibits stable characteristics. Hence, the observation sequence $\{r(k)\}$ is formed by adding the α–stable random noise $n(k)$ to the output of the channel $y(k)$ (see fig. 4.1)

$$r(k) = y(k) + n(k) \tag{4.5}$$

In symbol–by–symbol finite memory equalisers, the M most recent samples of the observation

sequence $\{r(k)\}$ are stored in the observation vector

$$\mathbf{r}(k) = [r(k) \quad r(k-1) \quad \ldots \quad r(k-M+1)]^{\mathrm{T}} \tag{4.6}$$

The information present in the observation M–vector $\mathbf{r}(k)$ is used by the equaliser to produce an estimate $\hat{x}(k-d)$ of the channel input $x(k-d)$. The equaliser often operates in two stages: first, a *decision function* $f_{\mathrm{d}}(\mathbf{r}(k))$ is evaluated on the received vector $\mathbf{r}$ producing a soft scalar output; then, a memoryless decision device selects the symbol from the transmitted alphabet that is closest to $f_{\mathrm{d}}(\mathbf{r}(k))$ (fig. 4.1). For binary data, the slicer is the sign function

$$\mathrm{sgn}(x) = \begin{cases} +1 & , \quad x \geq 0 \\ -1 & , \quad x < 0 \end{cases} \tag{4.7}$$

Such an equaliser is said to have order M and operate with lag or decision delay, d.

4.2 The optimum Bayesian equaliser

The most simple architecture in the class of equalisers making decisions in a symbol–by–symbol basis is the linear equaliser. In this case, $f_{\mathrm{d}}(\cdot)$ is restricted to be a linear function of $\mathbf{r}(k)$, and the task of the corresponding *linear equaliser* is to provide a causal approximation to the inverse of the channel, i.e. $H^{-1}(z)$. The output of the adaptive filter is then applied to the slicer to form an estimate of the transmitted sequence. Thus, linear equalisation is effectively an inverse filtering or deconvolution problem, and there are many linear adaptive filter algorithms available with which to train the equaliser [35]. However, the main disadvantage of the linear equaliser is that it may enhance the additive noise component [19]. Moreover, the adaptive filter does not exploit the fact that the transmitted sequence is drawn from a finite alphabet.

Indeed, the channel input vector $\mathbf{x}_{\mathrm{ch}}(k)$ contains N binary symbols, so that there are totally $N_{\mathrm{sc}} = 2^N$ possible discrete states for this vector, which we will denote as $\mathbf{x}_{\mathrm{ch}_i}$ ($0 \leq i < N_{\mathrm{sc}}$). But, since the channel is assumed to be a FIR filter, it follows from eq. (4.4) that the noiseless channel output $y(k)$ can only take N_{sc} possible discrete values, which we denote as $\bar{c}_i$ ($0 \leq i < N_{\mathrm{sc}}$). All these possible noiseless channel outputs are usually referred to as *scalar centers* and are given by

$$\bar{c}_i = \mathbf{h}^{\mathrm{T}} \mathbf{x}_{\mathrm{ch}_i} \quad , \quad 0 \leq i < N_{\mathrm{sc}} \tag{4.8}$$

The noiseless observation vector $\mathbf{y}(k)$, containing M sequential channel output observations, will exhibit discrete nature, as well. In order to explore the discrete character of the noise-free observation vector, it is convenient to describe the observation vector $\mathbf{r}(k)$ in matrix form [6, 76]. The state equation that relates the received samples vector $\mathbf{r}(k)$ to the vector of transmitted symbols $\mathbf{x}(k)$ is

$$\mathbf{r}(k) = \mathbf{H} \cdot \mathbf{x}(k) + \mathbf{n}(k) \tag{4.9}$$

where

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & h_2 & \cdots & h_{N-1} & 0 & 0 \\ 0 & h_0 & h_1 & \cdots & h_{N-2} & h_{N-1} & 0 \\ & \vdots & & & & \ddots & \\ 0 & 0 & 0 & \cdots & h_0 & \cdots & h_{N-1} \end{bmatrix} \tag{4.10}$$

is the $(M \times K)$ channel matrix, with $K = N + M - 1$. The vector $\mathbf{x}(k)$ is defined as

$$\mathbf{x}(k) = [x(k) \quad x(k-1) \quad \cdots \quad x(k-K+1)]^{\mathrm{T}} \tag{4.11}$$

and is the K–vector containing all the transmitted symbols that influence the observation vector. The vector

$$\mathbf{y}(k) = \mathbf{H} \cdot \mathbf{x}(k) \tag{4.12}$$

is the noise–free observation M–vector, that is, it contains the M most recent samples of the noise–free channel output sequence $\{y(k)\}$. But vector $\mathbf{x}(k)$ contains K binary symbols, therefore there are totally $N_{\mathrm{c}} = 2^K$ possible discrete states for this vector. We will denote all these N_{c} possible discrete states for the input vector $\mathbf{x}(k)$ as $\mathbf{x}_i$ $(0 \leq i < N_{\mathrm{c}})$. It follows from eq. (4.12) that the noise–free observation vector $\mathbf{y}(k)$ can take N_{c} discrete states, which we denote $\mathbf{c}_i$ $(0 \leq i < N_{\mathrm{c}})$. Hence,

$$\mathbf{c}_i = \mathbf{H} \cdot \mathbf{x}_i \quad , \quad 0 \leq i < N_{\mathrm{c}} \tag{4.13}$$

These discrete states $\mathbf{c}_i$ for the noise–free observation vector $\mathbf{y}(k)$ are called (vector) *centers*.

scalar centers	$\mathbf{x}_{\text{ch}}(k)$		$y(k)$
	$x(k)$	$x(k-1)$	
$\bar{c}_0$	-1	-1	-1.5
$\bar{c}_1$	-1	1	-0.5
$\bar{c}_2$	1	-1	0.5
$\bar{c}_3$	1	1	1.5

Table 4.1: *Scalar centers $\bar{c}_i$ for channel $H_1(z)$.*

centers	$\mathbf{x}(k)$			$\mathbf{y}(k)$		$\mathbf{c}_i$	
	$x(k)$	$x(k-1)$	$x(k-2)$	$y(k)$	$y(k-1)$	$\bar{c}_{l_0}$	$\bar{c}_{l_1}$
$\mathbf{c}_0$	-1	-1	-1	-1.5	-1.5	$\bar{c}_0$	$\bar{c}_0$
$\mathbf{c}_1$	-1	-1	1	-1.5	-0.5	$\bar{c}_0$	$\bar{c}_1$
$\mathbf{c}_2$	-1	1	-1	-0.5	0.5	$\bar{c}_1$	$\bar{c}_2$
$\mathbf{c}_3$	-1	1	1	-0.5	1.5	$\bar{c}_1$	$\bar{c}_3$
$\mathbf{c}_4$	1	-1	-1	0.5	-1.5	$\bar{c}_2$	$\bar{c}_0$
$\mathbf{c}_5$	1	-1	1	0.5	-0.5	$\bar{c}_2$	$\bar{c}_1$
$\mathbf{c}_6$	1	1	-1	1.5	0.5	$\bar{c}_3$	$\bar{c}_2$
$\mathbf{c}_7$	1	1	1	1.5	1.5	$\bar{c}_3$	$\bar{c}_3$

Table 4.2: *Discrete noise–free states of the output vector $\mathbf{y}(k)$ for $H_1(z)$.*

Finally, the vector $\mathbf{n}(k)$ contains the noise samples

$$\mathbf{n}(k) = [n(k) \quad n(k-1) \quad \cdots \quad n(k-M+1)]^{\text{T}} \tag{4.14}$$

For example, suppose the channel has two taps ($N = 2$) and a transfer function

$$H_1(z) = 1 + 0.5z^{-1} \tag{4.15}$$

We will examine a second order ($M = 2$) equaliser with decision lag, $d = 0$. There are totally $N_{\text{sc}} = 2^N = 4$ scalar centers for the channel noise free output $y(k)$, which are shown in table 4.1. The length of the transmitted symbols vector is $K = N + M - 1 = 3$. Therefore, there are a total of $N_{\text{c}} = 2^3 = 8$ states for $\mathbf{x}(k)$, and hence 8 (vector) centers for the noiseless channel output vector $\mathbf{y}(k)$. Table 4.2 shows all possible channel output vectors for this channel with 2 coefficients. It is worth noting that the vector centers $\mathbf{c}_i$ contain values that are taken from the

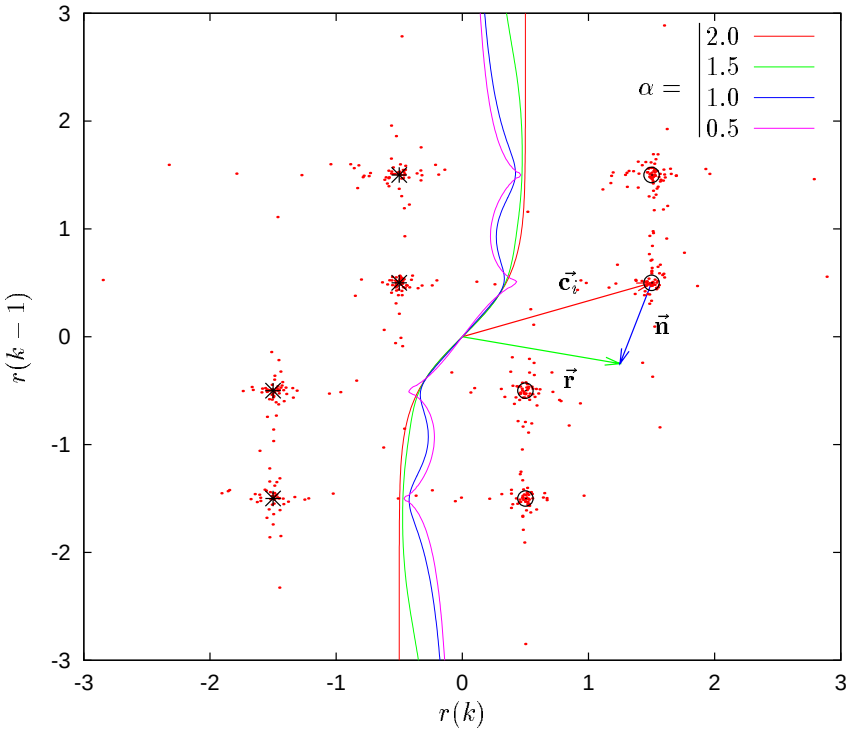


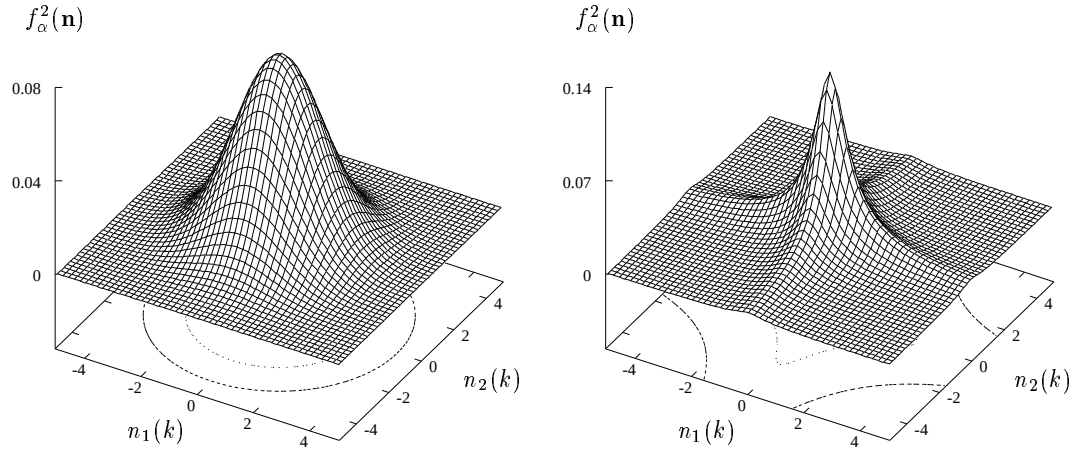
Figure 4.2: *Observation space and decision boundaries for the Bayesian equaliser for a variety of values for the characteristic exponent α. The channel is $H_1(z)$ ($\gamma = 0.1, M = 2, d = 0$).*

set of scalar centers $\bar{c}_i$. For example, $\mathbf{c}_2 = [\bar{c}_1, \quad \bar{c}_2]^{\mathrm{T}}$. In general, we can write the following relationship between the vector and scalar centers

$$\begin{aligned} \mathbf{c}_i &= [\bar{c}_{l_0}\ \bar{c}_{l_1}\ \ldots\ \bar{c}_{l_j}\ \ldots\ \bar{c}_{l_{M-1}}]^{\mathrm{T}} \\ l_j &= \left\lfloor \frac{i}{2^{M-1-j}} \right\rfloor \ \mathrm{rem}\ N_{\mathrm{sc}} \\ & \quad 0 \le i < N_{\mathrm{c}} \quad , \quad 0 \le j < M \end{aligned} \tag{4.16}$$

where rem denotes the remainder after the division. Figure 4.2 shows the observation space of the equaliser spanned by the received vector $\mathbf{r}(k)$. The noise–free centers are also depicted as either a $\divideontimes$ to indicate that the output vector represents an input $x(k-d) = -1$, or as a $\odot$ to represent an input $x(k-d) = +1$. As noise $\mathbf{n}(k)$ is added to the vector $\mathbf{y}(k)$, the 8 discrete centers become 8 clusters, as the points representing the received vector $\mathbf{r}(k)$ spread around the noise–free centers.

Given a center $\mathbf{c}_i$ has occurred, the vector $\mathbf{r}(k;\mathbf{c}_i)$ is an M–dimensional stable random variable with location $\mathbf{c}_i$. Therefore, the likelihood $p_{\mathbf{r}\mid\mathbf{c}}(\mathbf{r}(k) \mid \mathbf{c}_i)$ of the received vector $\mathbf{r}(k)$



(a) $\alpha = 2.0$ **(b)** $\alpha = 0.8$

Figure 4.3: *Likelihood of a two–dimensional α–stable random vector with location the null vector $\mathbf{0}$; non–Gaussian stable processes have radially asymmetric densities ($\gamma = 1$).*

conditioned on center $\mathbf{c}_i$ will be the M–dimensional multivariate α–stable distribution $f_\alpha^M(\cdot)$ centered at $\mathbf{c}_i$. Since the noise components of the received signal are assumed to be statistically independent, the simplified expression for the multivariate α–stable distribution may be used, given in eq. (3.36) (page 33). Specifically,

$$\begin{aligned} p_{\mathbf{r}\mid\mathbf{c}}(\mathbf{r}(k)\mid\mathbf{c}_i) &= f_\alpha^M(\mathbf{r}(k)-\mathbf{c}_i) \\ &= f_\alpha(\mathbf{r}_1(k)-\mathbf{c}_{i,1})\cdot f_\alpha(\mathbf{r}_2(k)-\mathbf{c}_{i,2})\cdots f_\alpha(\mathbf{r}_M(k)-\mathbf{c}_{i,M}) \end{aligned} \tag{4.17}$$

where $\mathbf{r}_j$ and $\mathbf{c}_{i,j}$ is the j–th element of vector $\mathbf{r}$ and $\mathbf{c}_i$, respectively.

Figure 4.3 shows the likelihood of a two–dimensional α–stable random vector

$$\mathbf{n}(k) = [n_1(k)\ n_2(k)]^{\mathrm{T}} \tag{4.18}$$

with location equal to the null vector $\mathbf{0}$. Here, $n_1(k)$ and $n_2(k)$ are independent α–stable random sequences. The cases of two different values of the characteristic exponent are depicted; a Gaussian ($\alpha = 2.0$ – fig. 4.3(a)) and a non–Gaussian ($\alpha = 0.8$ – fig. 4.3(b)). As the figure shows, the likelihood $p_{\mathbf{r}\mid\mathbf{c}}(\mathbf{r}(k)\mid\mathbf{c}_i)$ exhibits radial symmetry only in the $\alpha = 2$ case, resulting

from the interesting property of Gaussian densities [77]

$$\exp\left(-\frac{\mathbf{x}_1^2}{2\sigma^2}\right)\cdot\exp\left(-\frac{\mathbf{x}_2^2}{2\sigma^2}\right)\cdots\exp\left(-\frac{\mathbf{x}_M^2}{2\sigma^2}\right)=\exp\left(-\frac{\parallel\mathbf{x}\parallel^2}{2\sigma^2}\right) \tag{4.19}$$

where $\parallel\mathbf{x}\parallel$ is the *Euclidean norm* of vector $\mathbf{x}$.

Returning back to fig. 4.2, it becomes clear that the equalisation problem is one of classification. Specifically, the task is to partition the observation space spanned by vector $\mathbf{r}(k)$ into regions representing inputs of either $x(k-d)=-1$, or $x(k-d)=+1$. This leads directly to the Bayesian approach [5, 78].

The Bayesian or maximum *a posteriori* (MAP) symbol–by–symbol decision rule may be stated as follows: given the vector $\mathbf{r}(k)$, the symbol which maximises the posterior probabilities is chosen, i.e.,

$$\hat{x}(k-d)=\arg\max_{a\in A} P_{x\mid\mathbf{r}}\left(x(k-d)=a\mid\mathbf{r}(k)\right) \tag{4.20}$$

That is, having observed the vector $\mathbf{r}(k)$, we decide $+1$ if the probability that it was caused by $x(k-d)=+1$ exceeds the probability that it was caused by $x(k-d)=-1$, and vice versa.

The $N_{\mathrm{c}}=2^K$ discrete input vectors $\mathbf{x}_i$ ($0\leq iN_{\mathrm{c}}$) can be partitioned into two sets, conditioned on the transmitted symbol of interest:

$$S_{\mathbf{x}}^{-1}=[\mathbf{x}_i\ :\ x(k-d)=-1]\quad,\text{and}\quad S_{\mathbf{x}}^{+1}=[\mathbf{x}_i\ :\ x(k-d)=+1] \tag{4.21}$$

For example, $S_{\mathbf{x}}^{-1}$ contains all these combinations of binary symbols for which the d–th symbol is -1. The conditional probability of eq. (4.20) can be rewritten as the sum of all the conditional sequence probabilities associated with the input vectors of the appropriate subset $S_{\mathbf{x}}^{a}$, that is,

$$P_{\mathbf{x}\mid\mathbf{r}}(\mathbf{x}(k)=\mathbf{x}_i\mid\mathbf{r}(k))\quad,\quad\mathbf{x}_i\in S_{\mathbf{x}}^{a} \tag{4.22}$$

Thus, eq. (4.20) can be rewritten

$$\hat{x}(k-d)=\arg\max_{a\in A}\sum_{\{\mathbf{x}_i\in S_{\mathbf{x}}^{a}\}} P_{\mathbf{x}\mid\mathbf{r}}\left(\mathbf{x}(k)=\mathbf{x}_i\mid\mathbf{r}(k)\right) \tag{4.23}$$

Using Bayes' rule [16], the *a posteriori* probabilities may be expressed incorporating the

a priori probabilities

$$p(\mathbf{x} \mid \mathbf{r}) = \frac{p(\mathbf{r} \mid \mathbf{x})p(\mathbf{x})}{p(\mathbf{r})} \tag{4.24}$$

We have assumed in section 4.1 that the transmitted symbols are equiprobable; we also observe that the denominator in eq. (4.24) is independent of which symbol is transmitted. Consequently, the aforementioned decision rule is equivalent to finding the symbol $a \in A$ that maximises the likelihood of $\mathbf{r}(k)$ given $\mathbf{x}(k)$, that is $p(\mathbf{r} \mid x)$. Therefore, the decision rule of eq. (4.23) can be rewritten as follows

$$\hat{x}(k-d) = \arg\max_{a \in A} \sum_{\{\mathbf{x}_i \in S_{\mathbf{x}}^a\}} p_{\mathbf{r} \mid \mathbf{x}}\left(\mathbf{r}(k) \mid \mathbf{x}(k) = \mathbf{x}_i\right) \tag{4.25}$$

As eq. (4.12) implies, there is a one–to–one relationship between the possible input sequences $\mathbf{x}_i$ and the centers $\mathbf{c}_i$ of the channel. Therefore, we can accordingly partition the centers into two subsets, conditioned on each transmitted symbol of A:

$$S_{\mathrm{c}}^{-1} = [\mathbf{c}_i \ : \ x(k-d) = -1] \quad \text{, and} \quad S_{\mathrm{c}}^{+1} = [\mathbf{c}_i \ : \ x(k-d) = +1] \tag{4.26}$$

Thus, eq. (4.25) is equivalent to

$$\hat{x}(k-d) = \arg\max_{a \in A} \sum_{\{\mathbf{c}_i \in S_{\mathbf{c}}^a\}} p_{\mathbf{r} \mid \mathrm{c}}\left(\mathbf{r}(k) \mid \mathbf{y}(k) = \mathbf{c}_i\right) \tag{4.27}$$

Assuming a binary symmetric modulation constellation with equiprobable transmitted symbols, the decision rule of eq. (4.27) corresponds to a decision function

$$f_{\mathrm{d}}(\mathbf{r}(k)) = \sum_{\{\mathbf{c}_i \in S_{\mathrm{c}}^{+1}\}} p_{\mathbf{r} \mid \mathrm{c}}\left(\mathbf{r}(k) \mid \mathbf{y}(k) = \mathbf{c}_i\right) - \sum_{\{\mathbf{c}_i \in S_{\mathrm{c}}^{-1}\}} p_{\mathbf{r} \mid \mathrm{c}}\left(\mathbf{r}(k) \mid \mathbf{y}(k) = \mathbf{c}_i\right) \tag{4.28}$$

If we now define

$$s_i = \begin{cases} +1 & : \quad \mathbf{c}_i \ \in \ S_{\mathrm{c}}^{+1} \\ -1 & : \quad \mathbf{c}_i \ \in \ S_{\mathrm{c}}^{-1} \end{cases} \quad , \quad i = 0,\ 1,\ 2,\ \ldots\ N_{\mathrm{c}} \tag{4.29}$$

and use eq. (4.17) we obtain, as in [77],

$$\begin{aligned} f_{\mathrm{d}}(\mathbf{r}(k)) &= \sum_{i=0}^{N_{\mathrm{c}}-1} s_i \; p_{\mathbf{r} \mid \mathbf{c}} \left(\mathbf{r}(k) \mid \mathbf{y}(k) = \mathbf{c}_i \right) \\ &= \sum_{i=0}^{N_{\mathrm{c}}-1} s_i \; f_{\alpha}^{M}(\mathbf{r}(k) - \mathbf{c}_i) \end{aligned} \tag{4.30}$$

Equation eq. (4.30) implies that the M–dimensional α–stable distribution needs to be calculated on the term $\mathbf{r}(k) - \mathbf{c}_i$. To simplify the operation, replacing $f_{\alpha}^{M}(\cdot)$ from eq. (4.17) in eq. (4.30) we have

$$f_{\mathrm{d}}(\mathbf{r}(k)) = \sum_{i=0}^{N_{\mathrm{c}}-1} s_i \; \left\{ \prod_{j=1}^{M} f_{\alpha}(\mathbf{r}_j(k) - \mathbf{c}_{i,j}) \right\} \tag{4.31}$$

We can further replace $\mathbf{r}_j(k)$ with $r(k-j+1)$ and use eq. (4.16) to express eq. (4.31) in terms of the scalar centers

$$\begin{aligned} f_{\mathrm{d}}(\mathbf{r}(k)) &= \sum_{i=0}^{N_{\mathrm{c}}-1} s_i \; \left\{ \prod_{j=0}^{M-1} f_{\alpha}(r(k-j) - \bar{c}_{l_j}) \right\} \\ l_j &= \left\lfloor \frac{i}{2^{M-1-j}} \right\rfloor \; \mathrm{rem} \; N_{\mathrm{sc}} \end{aligned} \tag{4.32}$$

The alternative form of the decision function in eq. (4.32) is much easier to calculate since it involves scalar rather than vector centers (see also [79]). When the noise is Gaussian ($\alpha = 2$), the α–stable distribution $f_{\alpha}(\cdot)$ reduces to the normal density and therefore eq. (4.32) reduces to the traditional Bayesian equaliser.

Finally, the actual decision is taken upon the sign of f_{d}, that is

$$\hat{x}(k-d) = \mathrm{sgn}\left(f_{\mathrm{d}}(\mathbf{r}(k)) \right) \tag{4.33}$$

Equation (4.33) partitions the M–dimensional observation space spanned by the received signal vector $\mathbf{r}(k)$ in two sub–spaces. Therefore, the solution of equation

$$f_{\mathrm{d}}(\mathbf{r}(k)) = 0 \tag{4.34}$$

defines the optimum decision boundary. Since $f_{\mathrm{d}}(\mathbf{r}(k))$ is related to the distribution of the noise

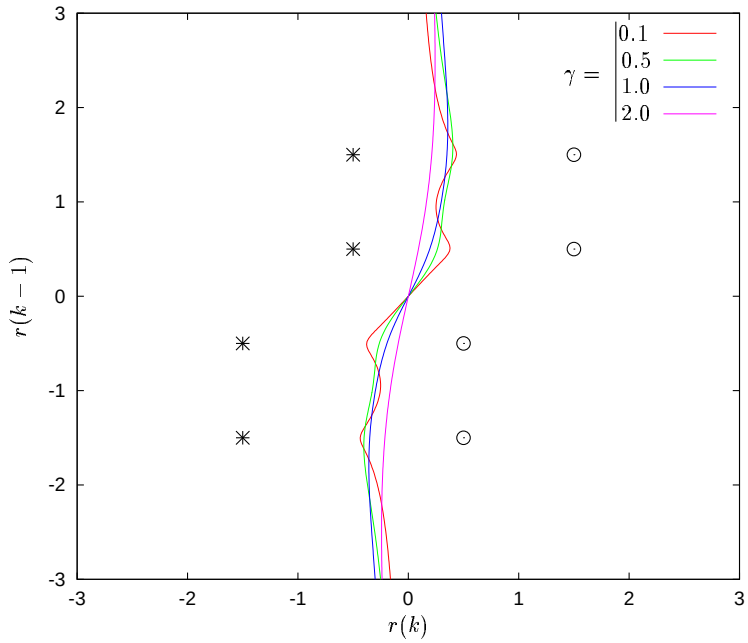


Figure 4.4: *The effects of the noise dispersion γ on the Bayesian decision boundaries for channel $H_1(z)$ ($\alpha = 1$, $M = 2$, $d = 0$).*

through eq. (4.30), the corresponding Bayesian decision boundaries will be inherently different for Gaussian and non–Gaussian noise distributions.

Consider again the minimum–phase channel channel studied in fig. 4.2 with response

$$H_1(z) = 1 + 0.5z^{-1} \tag{4.35}$$

and a Bayesian equaliser with order $M = 2$ and decision lag $d = 0$. Figure 4.2 also depicts how the Bayesian decision boundary changes as the noise distribution deviates from a pure Gaussian density ($\alpha = 2.0$) to a highly impulsive density ($\alpha = 1.5,\ 1.0,\ 0.5$). The dispersion of the noise is $\gamma = 0.1$ in all cases. It should be noted that a *traditional* Bayesian equaliser assumes that the interfering noise exhibits Gaussian characteristics; such an equaliser would produce the decision boundary corresponding to $\alpha = 2.0$. Consequently, in cases where the noise deviates from a pure Gaussian distribution, the discrepancies between the *optimum* Bayesian boundary and the boundary of a traditional Bayesian equaliser mean that the latter is no longer the optimum solution and that its performance will degrade; and the more the noise distribution deviates from the Gaussian density, the more the performance of a traditionally designed Bayesian equaliser degrades, in comparison to the optimum Bayesian equaliser.

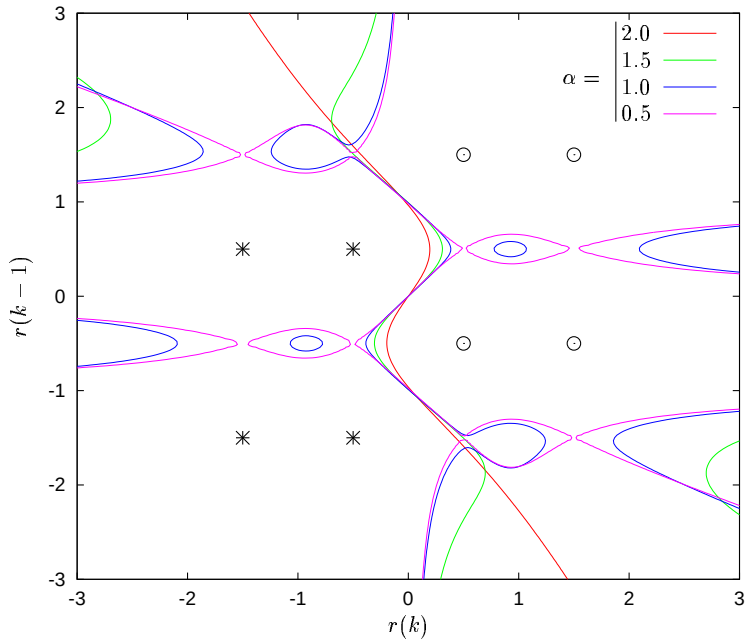


Figure 4.5: *Bayesian decision boundaries with channel $H_2(z)$ for a variety of values for the characteristic exponent α ($\gamma = 0.1$, $M = 2$, $d = 1$).*

The actual shape of the decision boundary defined by eq. (4.34) (with $f_{\rm d}(\mathbf{r}(k))$ defined in eq. (4.30)) depends highly on the noise dispersion γ, as well. Figure 4.4 depicts the optimum Bayesian boundary for different values of the noise dispersion, namely $\gamma = 0.1,\ 0.5,\ 1.0,\ 2.0$. The characteristic exponent is now fixed to $\alpha = 1.0$. For small values of γ the decision boundary is strongly affected by the characteristic exponent α, implying extended discrepancies between the traditional and optimum MAP equalisers. On the other hand, as γ increases the decision boundary becomes less dependent on α, resulting in very similar boundaries for both the traditional and the optimum Bayesian equalisers. That means, in turn, that the performance degradation of the traditional Bayesian equaliser in a non–Gaussian stable noise environment increases as the noise dispersion decreases [1].

Consider now the non–minimum phase channel having a transfer function with z transform

$$H_2(z) = 0.5 + z^{-1} \tag{4.36}$$

This is not a linearly separable channel; a decision lag $d = 1$ may be introduced so that the centers in $S_{\rm c}^{+1}$ can be linearly separated from those in $S_{\rm c}^{-1}$. In general, there is no decision

[1] It is worth pointing out that the use of signal–to–ratio (SNR) is avoided since it is meaningless for non–Gaussian stable noise.

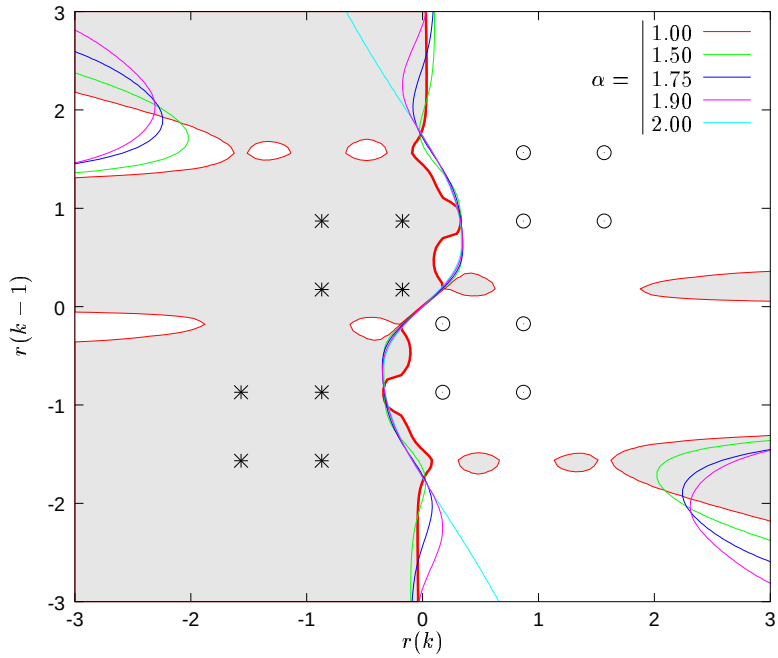


Figure 4.6: *Bayesian decision boundaries with channel $H_3(z)$ for different values of the characteristic exponent α ($\gamma = 0.1$, $M = 2$, $d = 1$). The shaded area depicts the -1 detection subspace for $\alpha = 1$.*

lag which can guarantee the linear separability of the corresponding centers. However, Chen *et al.* [80] proved a sufficient condition which guarantees the linear separability of the centers for the decision feedback equaliser which will be discussed in the following section. The Bayesian decision boundaries for channel $H_2(z)$ with different values of the characteristic exponent $\alpha = 2.0,\ 1.5,\ 1.0,\ 0.5$ is depicted in fig. 4.5. It is clear that the discrepancies between the traditional and the optimum Bayesian equaliser are significantly more extensive, compared with the study of channel $H_1(z)$.

Another example of non–minimum phase channel is the following:

$$H_3(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2} \tag{4.37}$$

Figure 4.6 depicts the decision boundaries implied by eq. (4.34) with this channel for Gaussian ($\alpha = 2.0$) and non–Gaussian ($\alpha = 1.90,\ 1.75,\ 1.50,\ 1.00$) stable noise. It is evident that when the noise deviates even slightly from the pure Gaussian distribution, the optimum Bayesian decision boundary changes strikingly, due to the lack of radial symmetry in the underlying noise probability density function.

A very important feature of the optimum decision boundaries for a non–Gaussian stable noise

distribution is the possible partitioning of the observation space in more than two contiguous subspaces. Specifically, as figures 4.5 and 4.6 demonstrate, the locus of points associated with a certain transmitted symbol may be the union of two or more disjoint sets. It is the heavy tails and the radial asymmetry of non–Gaussian stable densities, as well as the specific spatial arrangement of the centers, which produce disjoint classification sets in areas, where the traditional Bayesian decision function defines a contiguous classification subspace. Specifically, when the centers in the S_c^{-1} subset are not perfectly aligned (in the direction of the axes) with those in S_c^{+1}, the radially asymmetric pdf produces lobes of opposite sign in the opposite classification subspace.

A linear or a traditional Bayesian equaliser can reasonably approximate the *dominant* partitioning defined by the optimum decision boundary (depicted as a thick red line in fig. 4.6). However, they cannot model the disjoint areas of opposite sign. That implies that a linear or a traditional Bayesian equaliser will feature significantly degraded performance in non–Gaussian stable noise environments.

As suggested in [6, 32], a radial basis functions (RBF) network implementation of eq. (4.30) is a very natural choice when the noise distribution is Gaussian. However, as implied by fig. 4.3 and also discussed in [77], in non–Gaussian noise environments the basis functions are not radially symmetric. Hence the term *radial* is not appropriate but a very similar approach with RBF networks can be taken for the implementation of a non–Gaussian Bayesian equaliser defined by eq. (4.30).

4.3 Radial basis functions networks

The radial basis functions (RBF) network was originally developed for interpolation in multidimensional spaces [6, 81–83]. Consider a set of m–vectors $\{\boldsymbol{\rho}_i\}$ and a set of associated scalars $\{u_i\}$. The aim is to find a mapping $f : \mathbb{R}^m \to \mathbb{R}$ that satisfies

$$u_i = f(\boldsymbol{\rho}_i) \quad , \quad \forall i \tag{4.38}$$

The function $f(\boldsymbol{\rho})$ can then be used to interpolate the space $\mathbb{R}^m$ in all points $\boldsymbol{\rho} \in \mathbb{R}^m$. The schematic of a RBF network with m inputs and a scalar output is presented in fig. 4.7. This

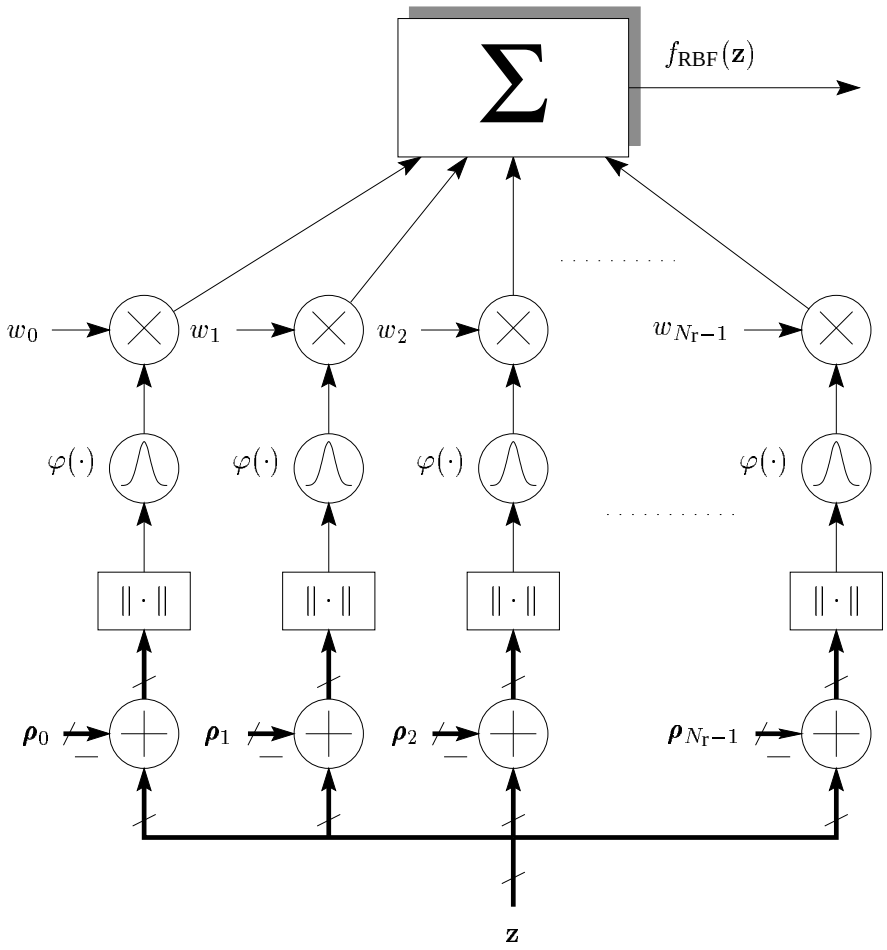


Figure 4.7: *A radial basis function network for multidimensional interpolation.*

network can implement a mapping $f_{\text{RBF}} : \mathbb{R}^m \to \mathbb{R}$, where $f_{\text{RBF}}(\cdot)$ is defined as

$$f_{\text{RBF}}(\mathbf{z}) = \sum_{i=0}^{N_\text{r}-1} w_i \, \varphi(\|\mathbf{z} - \boldsymbol{\rho}_i\|) \tag{4.39}$$

where $\mathbf{z} \in \mathbb{R}^m$ is the input vector, $\varphi(\cdot)$ is the *basis* function $\varphi : \mathbb{R}^+ \to \mathbb{R}$, $\boldsymbol{\rho}_i \in \mathbb{R}^m$ are known as RBF *centres* and w_i are the weights of the centers. That is, the distance of the input $\mathbf{z}$ to the corresponding center $\boldsymbol{\rho}_i$ is first extracted and then a scalar function $\varphi(\cdot)$ is calculated onto this distance. This operation is radially symmetric for an Euclidean vector norm $\|\cdot\|$, which gives rise to the name of this network.

Some common choices for the basis function $\varphi(\zeta)$ include a thin plate spline,

$$\varphi(\zeta) = \frac{\zeta}{\sigma_\text{r}^2} \log\left(\frac{\zeta}{\sigma_\text{r}}\right) \tag{4.40}$$

a multi quadratic,

$$\varphi(\zeta) = \sqrt{\zeta^2 + \sigma_{\mathrm{r}}^2} \tag{4.41}$$

an inverse multi–quadratic,

$$\varphi(\zeta) = \frac{1}{\sqrt{\zeta^2 + \sigma_{\mathrm{r}}^2}} \tag{4.42}$$

and Gaussian kernel,

$$\varphi(\zeta) = \exp\left(-\frac{\zeta^2}{2\sigma_{\mathrm{r}}^2}\right) \tag{4.43}$$

The parameter σ_{r}^2 controls the *radius* of influence of each basis function. The Gaussian and the inverse multi–quadratic kernel, in particular, are bounded and localised, in the sense that the basis functions decay to zero as $\zeta \to \infty$.

Broomhead and Lowe [84] reinterpreted the RBF network as a least square estimator which led to its wide use in signal processing applications such as time series prediction [85, 86], system identification [87, 88], interference cancellation [89], radar signal processing [90], pattern classification [91, 92] and channel equalisation [93]. Training of the RBF networks involves setting the parameters for the centres $\boldsymbol{\rho}_i$, radius σ_{r} and the linear weights w_i. The RBF networks are far easier to train compared to multilayer neural networks, since the training of centres, radius parameter and the weights can be done sequentially. The main characteristic of the RBF network is that it offers a nonlinear mapping, maintaining at the same time its linearity in parameter structure at the output layer.

In particular the RBF network provides a direct implementation of the Bayesian equaliser in a Gaussian noise environment. A deeper examination of the RBF decision function in eq. (4.39) in conjunction with a Gaussian kernel (eq. (4.43)), and the Bayesian equaliser decision function in eq. (4.30) (for $\alpha = 2$) shows that these functions are similar. Indeed, for $\alpha = 2$ the M–dimensional α–stable density reduces to the M–dimensional Gaussian density. The latter is essentially equivalent to the Gaussian basis function of eq. (4.43), because of the property of eq. (4.19). On the other hand, any scaled (through a positive multiplicative factor) form of the Bayesian decision function (eq. (4.30)) produces identical decision boundaries for binary trans-

mission, because of the slicer. Therefore, the RBF network can provide a Bayesian decision function for Gaussian noise environments by setting

- the RBF centres equal to the channel states $\boldsymbol{\rho}_i = \mathbf{c}_i$, $\forall i$,
- the RBF radius parameter equal to the Gaussian noise variance $\sigma_{\mathrm{r}}^2 = \sigma_n^2 \overset{\alpha=2}{=} 2\gamma$, and
- the linear weights

$$w_i = \begin{cases} +1 & : \quad \mathbf{c}_i \in S_{\mathbf{c}}^{+1} \\ -1 & : \quad \mathbf{c}_i \in S_{\mathbf{c}}^{-1} \end{cases}$$

This implementation of the Bayesian decision function is usually referred to as Bayesian RBF equaliser. The issues relating to the RBF equaliser design are discussed extensively in [6]. The RBF equalisers can provide optimal performance with small training sequences but they suffer from computational complexity. The number of RBF centres required in the equaliser increases exponentially with the equaliser order M and the channel delay dispersion order N, because the number of the centers is 2^{N+M-1}. This increases all the computations exponentially [31, 86].

In a non–Gaussian α–stable noise environment the M–dimensional density function $f_\alpha^M(\cdot)$ (eq. (4.30)) is not radially symmetric. Consequently, a scalar basis function $\varphi(\cdot)$ calculated on $||\mathbf{z} - \boldsymbol{\rho}_i||$ can not be used. It should be replaced by the actual M–dimensional noise distribution $f_\alpha^M(\cdot)$ calculated directly on the distance of the input vector $\mathbf{z}$ from center $\boldsymbol{\rho}_i$. Figure 4.8 depicts the functional diagram implementing the Bayesian equaliser for α–stable noise environments, with the equaliser notation of section 4.2 replacing the RBF notation. Since the basis function is not scalar the computational complexity is higher compared to the traditional Bayesian RBF for Gaussian noise. However, the alternative form of the decision function shown in eq. (4.32) can be used to reduce the overall complexity of calculations, as Patra *et al.* [94] suggested.

In early RBF equalisers [93] the RBF centres were selected at random, picked from a few of the initial input vectors. The weights were updated using supervised training by the LMS algorithm or its momentum version [95]. This resulted in equalisers with large number of centres making the network computationally complex. Chen proposed the OLS algorithm [85, 96] for selecting an optimum number of centres from a large number of candidate centres, resulting in near optimal performance. Subsequently, the close relationship between the RBF network and the Bayesian equaliser was found [76] and this provided the parametric implementation of the

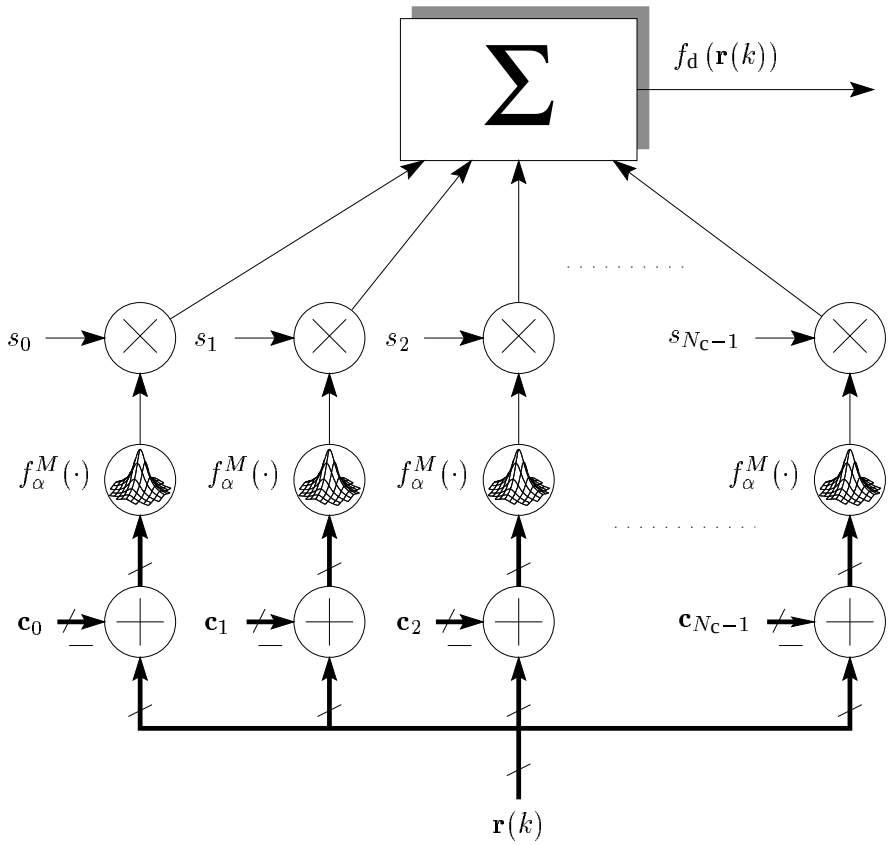


Figure 4.8: *Implementation of the Bayesian equaliser in α–stable noise.*

Bayesian equalisers with the RBF network. With the development of RBF networks that could handle complex signals [97], they were used for equalisation in communication systems with complex signal constellation [98]. Cha proposed the stochastic gradient algorithm [89] to adapt all the RBF parameters and used this technique to equalise 4–QAM digital communication systems.

One possible scheme employed for training the RBF centers in a supervised manner is to estimate the channel centres using a clustering algorithm (such as the κ–means clustering algorithm). Another approach is to first estimate the channel impulse response, then calculate the channel scalar centers from eq. (4.8), and finally the channel centers can be inferred from the channel scalar centers, through eq. (4.16). The latter scheme is more efficient, since the channel estimation is an integral operation in communication systems. Moreover, it requires a shorter training period to converge. However its use is restricted among scenarios where the channel is well modelled as a linear system. On the contrary, the direct clustering algorithm suffers from longer training periods increasing exponentially with respect to the channel and the equaliser

order. Nevertheless, it can cope with channels with non–linear impairements [31].

4.4 Decision feedback equaliser

In section 4.2 the optimum symbol–by–symbol maximum *a posteriori* probabilities (MAP) equaliser was derived. This equaliser is regarded as *feed–forward* because the decision function is computed on a set of past observations. However, it is possible to use a number of past decisions made by the equaliser to remove part or the whole of the intersymbol interference in order to enhance its performance. Recall from section 4.2 that all the transmitted symbols that influence the observation vector $\mathbf{r}(k)$ at time k are those contained in vector $\mathbf{x}(k)$

$$\mathbf{x}(k) = [x(k) \quad x(k-1) \quad \cdots \quad x(k-K+1)]^{\mathrm{T}} \tag{4.44}$$

We can initially assume that the D decisions $\{\hat{x}(k-L), \hat{x}(k-L-1) \ \ldots \ \hat{x}(k-K+1)\}$ (corresponding to the D oldest symbols in $\mathbf{x}(k)$) are correct, where $L = K - D$. Replacing these decisions [99–101] on the trailing part of the vector of transmitted symbols $\mathbf{x}(k)$ we have

$$\hat{\mathbf{x}}(k) = [x(k), \ \ldots \ , x(k-L+1) \quad | \quad \hat{x}(k-L), \ \ldots \ , \hat{x}(k-K+1)]^{\mathrm{T}}$$

That is, L is the length of the heading part of $\hat{\mathbf{x}}(k)$ containing the actual transmitted symbols, which we define as the subvector

$$\mathbf{x}_{\mathrm{R}}(k) = [x(k), \ x(k-1), \ \ldots \ , \ x(k-L+1)]^{\mathrm{T}}$$

The length of the trailing part of $\hat{\mathbf{x}}(k)$ which contains past detected symbols is D and defines the subvector [2]

$$\hat{\mathbf{x}}_{\mathrm{D}}(k) = [\hat{x}(k-L), \ \hat{x}(k-L-1), \ \ldots \ , \ \hat{x}(k-K+1)]^{\mathrm{T}}$$

Hence, vector $\hat{\mathbf{x}}(k)$ can be rewritten in a compact form as follows

$$\hat{\mathbf{x}}(k) = \left[\frac{\mathbf{x}_{\mathrm{R}}(k)}{\hat{\mathbf{x}}_{\mathrm{D}}(k)} \right]$$

[2] The subscript R stands for *residual* while D stands for *feedback*

The number D is the *feedback order* of the equaliser and can take integer values $0 \leq D < N + M - 1$. The right hand side inequality means that the residual input vector $\mathbf{x}_{\mathrm{R}}(k)$ should have at least one element. Moreover, there is a constraint for the decision lag of the equaliser $d < K - D$, which ensures that the detected symbols contained in the feedback input vector $\mathbf{x}_{\mathrm{D}}$ have previously been processed by the equaliser. In other words, the most recent symbol in $\mathbf{x}_{\mathrm{D}}$ can only be $\hat{x}(k - d - 1)$.

If we appropriately partition the channel matrix as

$$\mathbf{H} = \left[\begin{array}{cccc|cccccc} h_0 & h_1 & \cdots & h_{L-1} & h_L & \cdots & h_{N-1} & 0 & \cdots & 0 \\ 0 & h_0 & \cdots & h_L & h_{L+1} & \cdots & h_{N-2} & h_{N-1} & \cdots & 0 \\ & \vdots & & & & & \ddots & & & \\ 0 & 0 & \cdots & & & & & \cdots & & h_{N-1} \end{array}\right] \tag{4.45}$$

the following sub–matrices are defined; the residual part $\mathbf{H}_{\mathrm{R}}$ with size $(M \times L)$

$$\mathbf{H}_{\mathrm{R}} = \left[\begin{array}{cccc} h_0 & h_1 & \cdots & h_{L-1} \\ 0 & h_0 & \cdots & h_L \\ & \vdots & & \\ 0 & 0 & \cdots & \end{array}\right] \tag{4.46}$$

and the feedback part $\mathbf{H}_{\mathrm{D}}$ with size $(M \times D)$

$$\mathbf{H}_{\mathrm{D}} = \left[\begin{array}{cccccc} h_L & \cdots & h_{N-1} & 0 & \cdots & 0 \\ h_{L+1} & \cdots & h_{N-2} & h_{N-1} & \cdots & 0 \\ & & \ddots & & & \\ & & & \cdots & & h_{N-1} \end{array}\right] \tag{4.47}$$

Using these definitions, we can rewrite the state equation that relates the observation vector $\mathbf{r}(k)$ with the transmitted symbols vector $\mathbf{x}(k)$ (eq. (4.9)) as [6]

$$\mathbf{r}(k) = \left[\begin{array}{c|c} \mathbf{H}_{\mathrm{R}} & \mathbf{H}_{\mathrm{D}} \end{array}\right] \left[\begin{array}{c} \mathbf{x}_{\mathrm{R}}(k) \\ \hline \mathbf{x}_{\mathrm{D}}(k) \end{array}\right] + \mathbf{n}(k) \tag{4.48}$$

The effect of the decisions contained in $\mathbf{x}_{\mathrm{D}}(k)$ can then be removed from the observation vector

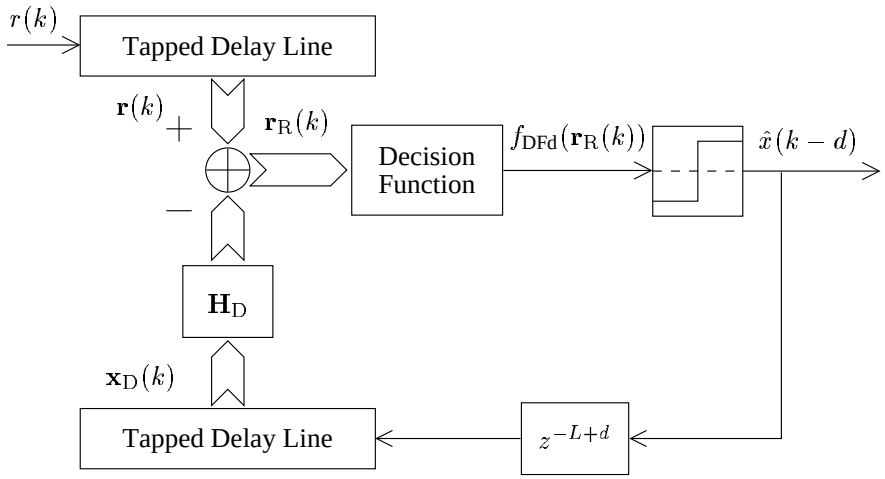


Figure 4.9: *Decision feedback equaliser.*

$\mathbf{r}(k)$ to produce a new residual observation vector, defined as

$$\begin{aligned} \mathbf{r}_{\mathrm{R}}(k) &\stackrel{\triangle}{=} \mathbf{r}(k) - \mathbf{H}_{\mathrm{D}}\mathbf{x}_{\mathrm{D}}(k) \\ &= \mathbf{H}_{\mathrm{R}}\mathbf{x}_{\mathrm{R}}(k) + \mathbf{n}(k) \\ &= \mathbf{y}_{\mathrm{R}}(k) + \mathbf{n}(k) \end{aligned} \tag{4.49}$$

where

$$\mathbf{y}_{\mathrm{R}}(k) = \mathbf{H}_{\mathrm{R}}\mathbf{x}_{\mathrm{R}}(k) \tag{4.50}$$

is the residual channel output vector. We can now apply a Bayesian decision function to the residual observation vector $\mathbf{r}_{\mathrm{R}}(k)$ rather than $\mathbf{r}(k)$. A decision feedback equaliser (DFE) with feedback order D implementing this scheme is depicted in fig. 4.9.

There are $N_{\mathrm{DFc}} = 2^L$ possible discrete states $\mathbf{c}_i^{\mathrm{R}}$ (centers of the DFE) for vector $\mathbf{y}_{\mathrm{R}}(k)$ which can be computed by the formula

$$\mathbf{c}_i^{R} = \mathbf{H}_{\mathrm{R}}\mathbf{x}_{\mathrm{R}i} \quad , \quad 0 \leq i < N_{\mathrm{DFc}} \tag{4.51}$$

where $\mathbf{x}_{\mathrm{R}i}$ are all N_{DFc} possible discrete states for the residual input vector $\mathbf{x}_{\mathrm{R}}(k)$. Since the number of centers for the feed–forward Bayesian equaliser is $2^K \equiv 2^{L+D}$, this results in a computational complexity reduction of the order of 2^D. As in section 4.2, the DFE centers $\mathbf{c}_i^{\mathrm{R}}$

can be partitioned into two subsets, conditioned on the transmitted symbol of interest

$$S_{\mathrm{c^R}}^{-1} = \left[\mathbf{c}_i^{\mathrm{R}} \ : \ x(k-d) = -1\right] \quad \text{, and} \quad S_{\mathrm{c^R}}^{+1} = \left[\mathbf{c}_i^{\mathrm{R}} \ : \ x(k-d) = +1\right] \tag{4.52}$$

Finally, the appropriate sign for each center should be defined as follows

$$s_{\mathrm{DF}i} = \begin{cases} +1 & : \quad \mathbf{c}_i^{\mathrm{R}} \in S_{\mathrm{c^R}}^{+1} \\ -1 & : \quad \mathbf{c}_i^{\mathrm{R}} \in S_{\mathrm{c^R}}^{-1} \end{cases} \quad , \quad i = 0,\ 1,\ 2,\ \ldots\ N_{\mathrm{DFc}} \tag{4.53}$$

Therefore, the Bayesian decision function with decision feedback takes the form

$$f_{\mathrm{DFd}}(\mathbf{r}_{\mathrm{R}}(k)) \quad = \quad \sum_{i=0}^{N_{\mathrm{DFc}}-1} s_{\mathrm{DF}i}\ f_{\alpha}^{M}(\mathbf{r}_{\mathrm{R}}(k) - \mathbf{c}_i^{\mathrm{R}}) \tag{4.54}$$

As an example, consider the channel with transfer function

$$H_3(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2} \tag{4.55}$$

and a decision feedback equaliser with feed–forward order $M = 2$, feedback order $D = 2$ and decision lag $d = 1$. The length of the input vector $\mathbf{x}$ is $K = N + M - 1 = 4$ and the length of the residual input vector $\mathbf{x}_{\mathrm{R}}$ is $L = K - D = 2$. Then, the (2×2) residual channel matrix $\mathbf{H}_{\mathrm{R}}$ and the (2×2) feedback channel matrix $\mathbf{H}_{\mathrm{D}}$ take the form

$$\mathbf{H}_{\mathrm{R}} = \begin{bmatrix} 0.3482 & 0.8704 \\ 0 & 0.3482 \end{bmatrix} \quad , \quad \mathbf{H}_{\mathrm{D}} = \begin{bmatrix} 0.3482 & 0 \\ 0.8704 & 0.3482 \end{bmatrix} \tag{4.56}$$

We can now construct a (2×4) matrix $\mathbf{X}_{\mathrm{R}}$ containing all possible combinations of symbols for the residual input vector $\mathbf{x}_{\mathrm{R}}(k)$ as follows

$$\mathbf{X}_{\mathrm{R}} = \begin{bmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \end{bmatrix} \tag{4.57}$$

Multiplying $\mathbf{H}_{\mathrm{R}}$ with $\mathbf{X}_{\mathrm{R}}$ we obtain all 4 centers of the decision feedback equaliser in matrix

centers	$\mathbf{x}_{\mathrm{R}}(k)$		$\mathbf{y}_{\mathrm{R}}(k)$	
	$x_{\mathrm{R}}(k)$	$x_{\mathrm{R}}(k-1)$	$y_{\mathrm{R}}(k)$	$y_{\mathrm{R}}(k-1)$
$\mathbf{c}_0^{\mathrm{R}}$	-1	-1	-1.2186	-0.3482
$\mathbf{c}_1^{\mathrm{R}}$	-1	1	0.5222	0.3482
$\mathbf{c}_2^{\mathrm{R}}$	1	-1	-0.5222	-0.3482
$\mathbf{c}_3^{\mathrm{R}}$	1	1	1.2186	0.3482

Table 4.3: *Discrete noise–free states of the residual output vector* $\mathbf{y}^R(k)$ *for* $H_3(z)$.

$\mathbf{C}_{\mathrm{R}}$

$$\mathbf{C}_{\mathrm{R}} = \mathbf{H}_{\mathrm{R}}\mathbf{X}_{\mathrm{R}} = \begin{bmatrix} -1.2186 & 0.5222 & -0.5222 & 1.2186 \\ -0.3482 & 0.3482 & -0.3482 & 0.3482 \end{bmatrix} \tag{4.58}$$

which are summarised in table 4.3.

The optimum decision boundaries of the Bayesian feed–forward equaliser for this channel were depicted in fig. 4.6 (page 53). Figure 4.10 depicts the corresponding boundaries for the Bayesian DFE. The features of the optimum decision boundaries are again significantly different compared to the boundaries of a traditionally (under the Gaussian assumption) designed equaliser. Therefore, it is reasonable to expect a considerable performance degradation of the traditional Bayesian equaliser in a non–Gaussian noise environment.

In practice, of course, the symbols contained in the feedback vector $\mathbf{x}_{\mathrm{D}}(k)$ may not be identical to the corresponding transmitted data, due to occasional errors in the detection. This discrepancy can cause further errors, an effect usually called *error propagation*. However, as the experimental data show, the performance degradation due to error propagation is low compared to the advantage in performance due to the adoption of the decision feedback scheme.

A very desirable characteristic of the decision feedback equaliser is that, under certain conditions, it can ensure the linear separability of the subset of the channel centers. Chen *et al.* have shown in [80] that a sufficient condition is to set the decision lag $d = N - 1$, the equaliser order $M = N$ and the feedback order $D = N - 1$. These values result in a residual channel matrix $\mathbf{H}_{\mathrm{R}}$ which is square upper triangular Töplitz. Therefore, its eigenvalues are its diagonal elements which are all h_0. Thus, $\mathbf{H}_{\mathrm{R}}$ is full rank and its inverse always exists. The details of

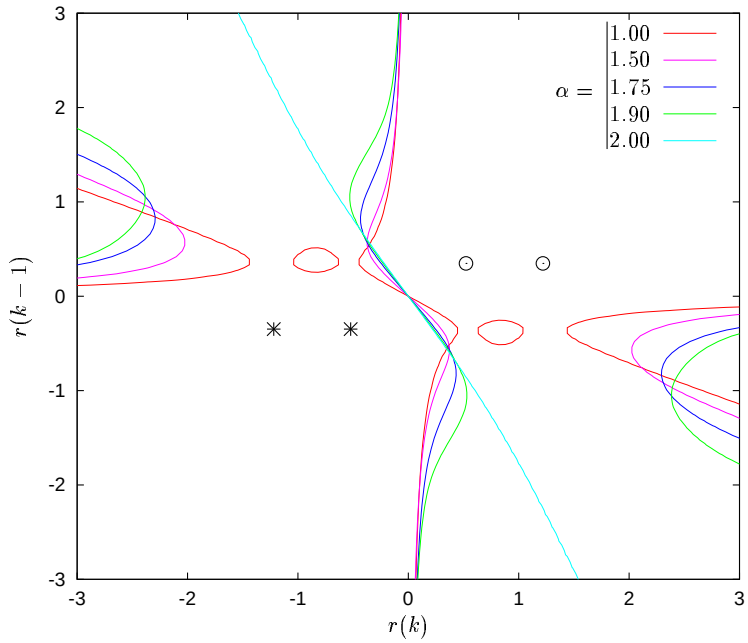


Figure 4.10: *Decision boundaries of the Bayesian DFE with channel $H_3(z)$ for a variety of values for the characteristic exponent α ($\gamma = 0.1$, $M = 2$, $D = 2$, and $d = 1$).*

the proof can be found in [80].

4.5 Evaluating systems in infinite power noise environment

The traditional performance measures are usually plots of the bit–error ratio (BER) against the signal–to–noise ratio (SNR). In a Gaussian noise environment the computation of SNR involves the variance of the useful received signal, as well as the variance of the corrupting noise. In non–Gaussian stable noise environment (α–stable noise with $\alpha < 2$), however, the variance of the noise is infinite [9], making the use of traditional BER–to–SNR graphs meaningless.

Nevertheless, all receivers in practice have a finite input dynamic range. Consider the generic receiver depicted in fig. 4.11. The limiter at the front end of the receiver is assumed to be an ideal saturation device, with transfer function

$$g(x, G) = \begin{cases} x & : \quad |x| \leq G \\ \operatorname{sgn}(x)\ G & : \quad \text{elsewhere} \end{cases} \tag{4.59}$$

G being the saturation point of the limiter. For a given saturation limit G, the SNR at the limited received signal $r_{\mathrm{L}}(k)$ is always finite. In this thesis we propose that the SNR of the limited

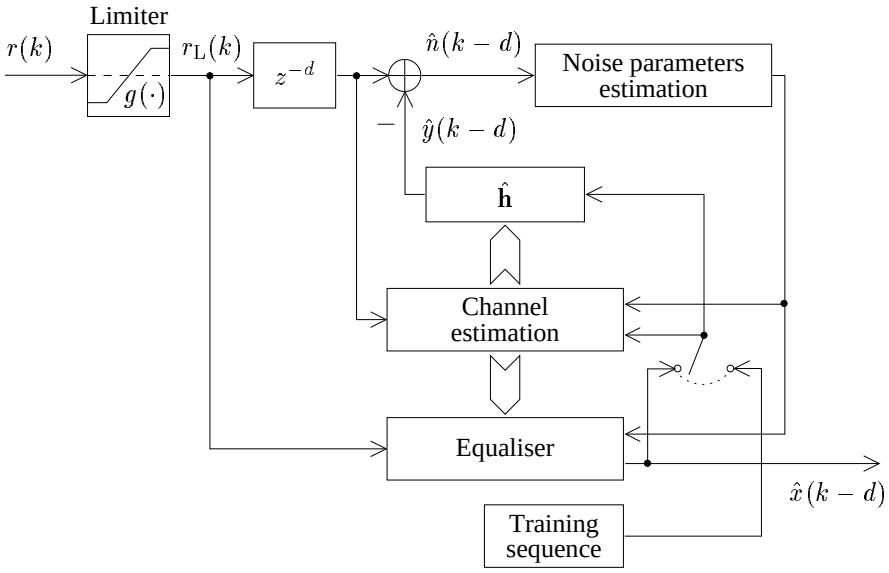


Figure 4.11: *Generic adaptive equaliser with saturation device at the front end.*

received signal $r_{\mathrm{L}}(k)$ should be used for performance evaluation in environments where the noise variance is infinite. We will refer to this as the *SNR at the receiver*. In the following we present some analytical tools that enable us to calculate this signal–to–noise ratio.

The distribution of the received signal $r(k)$ is

$$f_r(s) = \frac{1}{N_{\mathrm{sc}}} \sum_{i=1}^{N_{\mathrm{sc}}} f_\alpha(s - \bar{c}_i) \tag{4.60}$$

where $N_{\mathrm{sc}} = 2^N$ is the number of the scalar centers $\bar{c}_i$ of the channel (see eq. (4.8)). The limiter g truncates the pdf of the received signal and its tails are concentrated at the points $+G, -G$ where they appear as *Dirac* impulses $\delta(s)$ (fig. 4.13). The pdf of the limited received sequence $r_{\mathrm{L}}(k)$ is therefore

$$f_{r_{\mathrm{L}}}(s) = \frac{1}{N_{\mathrm{sc}}} \sum_{i=1}^{N_{\mathrm{sc}}} \bar{f}_\alpha(s - \bar{c}_i, -G - \bar{c}_i, G - \bar{c}_i) \tag{4.61}$$

where $\bar{f}_\alpha(s, G_1, G_2)$ is the α–stable pdf *truncated* at the points G_1 and G_2 and is given by

$$\bar{f}_\alpha(s, G_1, G_2) = f_\alpha(s)\,\Pi(s, G_1, G_2) + I_{\mathrm{l}}(G_1)\delta(s - G_1) + I_{\mathrm{r}}(G_2)\delta(s - G_2)$$

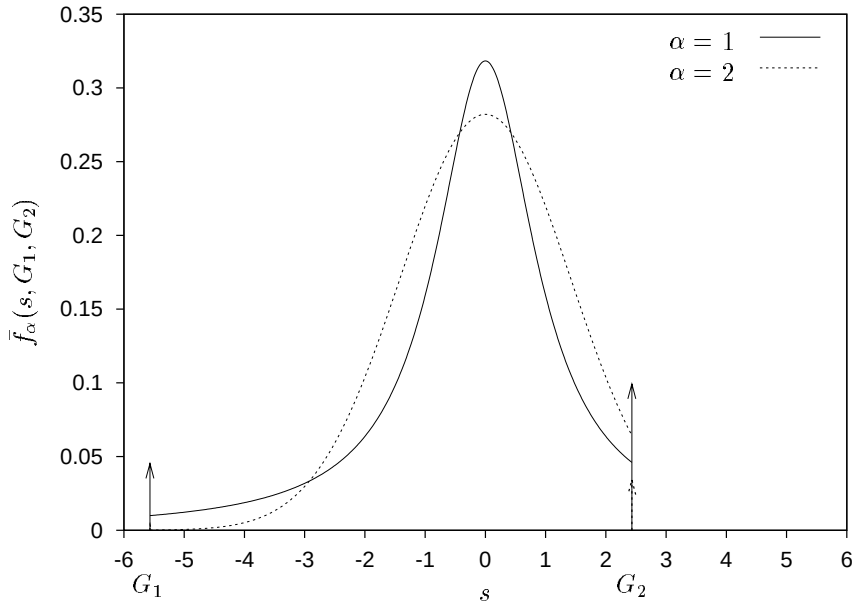


Figure 4.12: *The truncated α–stable distribution at points $G_1 < G_2$ for the Gaussian ($\alpha = 2$) and Cauchy ($\alpha = 1$) distributions ($\gamma = 1, G_1 = -5.6, G_2 = 2.4$).*

where

$$\begin{aligned} \Pi(s, G_1, G_2) &= \begin{cases} 1 & : \quad G_1 < s < G_2 \\ 0 & : \quad \text{elsewhere} \end{cases} \\ I_{\mathrm{l}}(G) &= \int_{-\infty}^{G} f_\alpha(s)\, ds \\ I_{\mathrm{r}}(G) &= \int_{G}^{\infty} f_\alpha(s)\, ds \end{aligned} \tag{4.62}$$

Figure 4.12 pictures the truncated α–stable distribution for two different values of the characteristic exponent. The $\delta(s)$ pulses at points G_1 and G_2 are depicted as arrows whose size are proportional to the pulse relative height. As an example for the distribution of the limited received signal we consider the channel with transfer function

$$H_3(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2} \tag{4.63}$$

The scalar centers of this channel are $2^3 = 8$ and are depicted in the table 4.4. Suppose the limiter at the front end of the receiver has a dynamic range $G = 2.2$. Figure 4.13 depicts the distribution of the limited received signal $r_{\mathrm{L}}(k)$ for two values of the noise characteristic exponent. It is worth noting that two pairs from the scalar centers are identical, so the peaks of

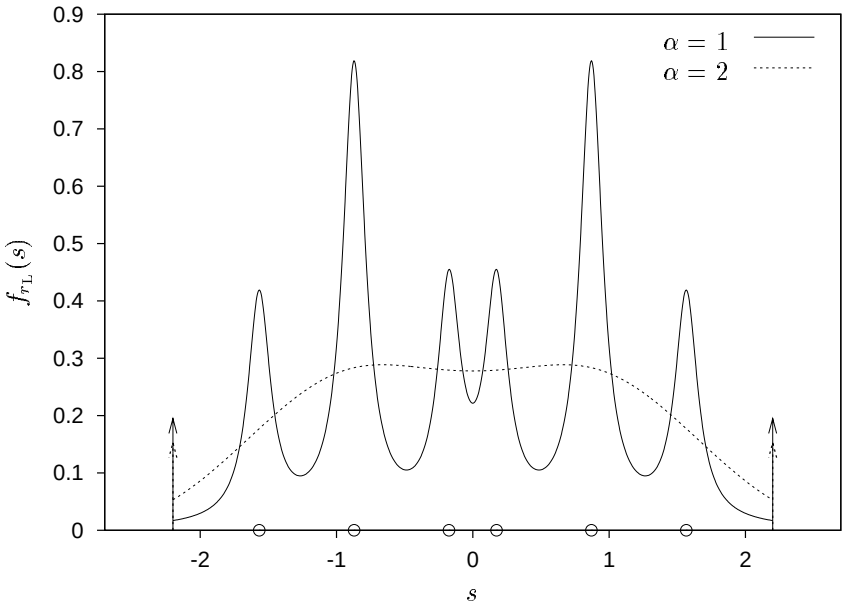


Figure 4.13: *The pdf of the limited received signal $r_{\mathrm{L}}(k)$ for Gaussian ($\alpha = 2$, $\gamma = 0.135$) and impulsive noise ($\alpha = 1$, $\gamma = 0.1$) with channel $H_3(z)$. The limiting level is $G = 2.2$.*

the pdf at these points are double the height of the rest. The power from the tails of the noise distribution concentrates at -2.2 and $+2.2$ and appears as Dirac impulses. The circles $\circ$ denote the corresponding scalar centers.

$\bar{c}_0$	$\bar{c}_1$	$\bar{c}_2$	$\bar{c}_3$	$\bar{c}_4$	$\bar{c}_5$	$\bar{c}_6$	$\bar{c}_7$
-1.5668	-0.8704	0.1740	0.8704	-0.8704	-0.1740	0.8704	1.5668

Table 4.4: *The scalar centers $\bar{c}_i$ of channel $H_3(z)$.*

The receiver removes the channel output estimate $\hat{y}(k-d)$ from the limited received signal $r_{\mathrm{L}}(k-d)$ to form an estimate of the noise samples $\hat{n}(k-d)$ (fig. 4.11). We can assume, without loss of generality, that the estimates $\hat{y}(k)$ are correct. The pdf of the noise estimate $\hat{n}(k)$ will then be

$$f_{\hat{n}}(s) = \frac{1}{N_{\mathrm{sc}}} \sum_{i=1}^{N_{\mathrm{sc}}} \bar{f}_\alpha(s, -G - \bar{c}_i, G - \bar{c}_i) \tag{4.64}$$

Figure 4.14 depicts an example for the distribution of the noise estimate sequence $\hat{n}(k)$. The dynamic range of the limiter is now $G = 4$. For Gaussian noise ($\alpha = 2$) the dispersion is set to $\gamma = 4.07$ and for Cauchy noise ($\alpha = 1$) the dispersion is $\gamma = 1.5$. These values for the

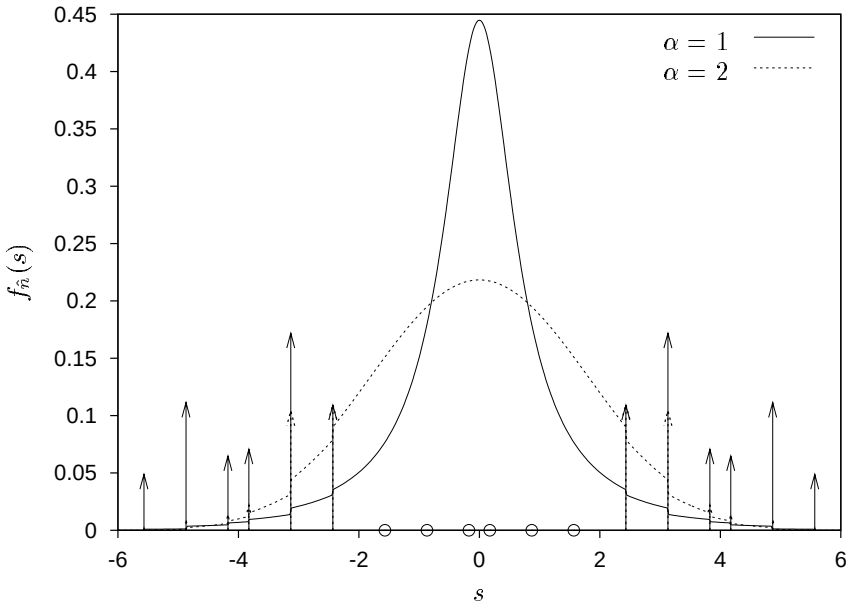


Figure 4.14: *The pdf of the noise estimate $\hat{n}(k)$ for Gaussian ($\alpha = 2, \gamma = 1.67$) and impulsive noise ($\alpha = 1, \gamma = 0.72$) for $G = 4$. The channel is $H_3(z)$.*

dispersion result in the same variance for the noise estimate signal. In that sense, a traditional Bayesian equaliser, assuming that the noise is Gaussian, would set the radius parameter σ_{r}^2 (see eq. (4.43)) equal to the variance of the noise estimate $\hat{n}(k)$. Hence, a traditional Bayesian equaliser would falsely assume that the noise estimate distribution is given by the dashed line in fig. 4.14 rather than the correct solid line. It is noteworthy that the choice for the dispersion values in fig. 4.13 also results in the same variance for the noise estimate.

Due to the symmetry of scalar centers, $f_{\hat{n}}(s)$ is symmetric. Therefore, the mean of $\hat{n}(k)$ is zero, while its variance can be written

$$v_{\hat{n}}(\alpha, \gamma, G) = \int_{-\infty}^{\infty} s^2 \, f_{\hat{n}}(s) \, ds = \frac{1}{N_{\mathrm{sc}}} \sum_{i=1}^{N_{\mathrm{sc}}} \int_{-\infty}^{\infty} s^2 \, \bar{f}_{\alpha}(s, -G - \bar{c}_i, G - \bar{c}_i) ds \tag{4.65}$$

The integral at the rightmost part of eq. (4.65) can further be expressed as

$$\begin{aligned} V(\alpha, \gamma, G_1, G_2) &= \int_{-\infty}^{\infty} s^2 \, \bar{f}_{\alpha}(s, G_1, G_2) ds \\ &= \int_{-\infty}^{\infty} s^2 \left\{ f_{\alpha}(s) \, \Pi(s, G_1, G_2) + I_{\mathrm{l}}(G_1) \delta(s - G_1) + I_{\mathrm{r}}(G_2) \delta(s - G_2) \right\} ds \\ &= G_1^2 \, I_{\mathrm{l}}(G_1) + G_2^2 \, I_{\mathrm{r}}(G_2) + \int_{G1}^{G2} s^2 \, f_{\alpha}(s) ds \end{aligned} \tag{4.66}$$

In general, $f_\alpha(s)$ cannot be expressed in closed form except the $\alpha = 2$ and $\alpha = 1$ cases. For these two special cases, it is possible to calculate $V(\alpha, \gamma, G_1, G_2)$. For the Gaussian case ($\alpha = 2$) we obtain

$$\begin{aligned} V(2,\gamma,G_1,G_2) &= \frac{G_1^2+G_2^2}{2} + \frac{1}{2}\,(2\,\gamma - G_2^2)\,\mathrm{erf}\left(\frac{G_2}{2\,\sqrt{\gamma}}\right) \\ &\quad - \frac{1}{2}\,(2\,\gamma - G_1^2)\,\mathrm{erf}\left(\frac{G_1}{2\,\sqrt{\gamma}}\right) \\ &\quad - 2\,\sqrt{\frac{\gamma}{\pi}}\,\left(G_2\,\exp\left(-\frac{G_2^2}{4\,\gamma}\right) + G_1\,\exp\left(-\frac{G_1^2}{4\,\gamma}\right)\right) \end{aligned} \tag{4.67}$$

where $\mathrm{erf}(\cdot)$ is the *error function*, defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t^2}\,dt \tag{4.68}$$

Accordingly, for the Cauchy case ($\alpha = 1$)

$$\begin{aligned} V(1,\gamma,G_1,G_2) &= \frac{G_1^2+G_2^2}{2} + \frac{\gamma\,(G_2-G_1)}{\pi} \\ &\quad - \frac{(\gamma^2+G_2^2)\mathrm{atan}\left(\frac{G_2}{\gamma}\right) - (\gamma^2+G_1^2)\mathrm{atan}\left(\frac{G_1}{\gamma}\right)}{\pi} \end{aligned} \tag{4.69}$$

From eq. (4.65) we can write the variance of the noise estimate $\hat{n}(k)$ as

$$v_{\hat{n}}(\alpha,\gamma,G) = \frac{1}{N_{\mathrm{sc}}} \sum_{i=1}^{N_{\mathrm{sc}}} V(\alpha,\gamma,-G-\bar{c}_i,G-\bar{c}_i) \tag{4.70}$$

The signal–to–noise ratio *at the receiver* (in dB) can now be expressed as a function of the noise parameters α, γ and the dynamic range of the receiver G

$$\mathrm{SNR}_{\mathrm{rcv}} = 10\,\log_{10}\left(\frac{v_y}{v_{\hat{n}}(\alpha,\gamma,G)}\right) \tag{4.71}$$

where v_y is the variance of the noise–free channel output.

In practice, for a given $\mathrm{SNR}_{\mathrm{rcv}}$, characteristic exponent α and dynamic range of the receiver G, it is possible to numerically solve eq. (4.71) for the noise dispersion γ. For values of α that it is not possible to analytically compute eq. (4.71) the actual variance of the noise estimate $\hat{n}(k)$ may be experimentally measured in order to compute the working SNR. However, in

section 6.4 we suggest an approximate method to compute the variance $v_{\hat{n}}(\alpha, \gamma, G)$ for a given dispersion γ. Accordingly, using an analogous approximation γ can be obtained for a given $\mathrm{SNR}_{\mathrm{rcv}}$.

4.6 Experiments

In order to assess the Bayesian equaliser in an α–stable noise environment, the experimental performance of a number of feed–forward and decision feedback equalisers was recorded. The simulations were performed for a variety of values for the noise characteristic exponent, ranging from pure Gaussian to highly impulsive noise environments. As mentioned earlier, for Gaussian noise the optimum equaliser is identical to the traditionally designed one. The *optimum* equalisers (designed for the actual α–stable noise distribution) are compared with the corresponding traditionally designed Bayesian equaliser (which assumes that the noise statistics follow Gaussian properties).

The optimum equalisers were provided with the correct parameters of the noise distribution, that is the characteristic exponent α and the dispersion parameter γ. For the traditionally designed equaliser the radius parameter σ_{r}^2 was set equal to the variance of the noise estimate at the receiver $\hat{n}(k)$. For this set of experiments the equalisers had perfect knowledge of the channel impulse response.

The simulations have been carried out for three different channels, having the following transfer functions

$$\begin{aligned} H_1(z) &= 1 + 0.5z^{-1} \\ H_2(z) &= 0.5 + z^{-1} \\ H_3(z) &= 0.3482 + 0.8704z^{-1} + 0.3482z^{-2} \end{aligned} \tag{4.72}$$

The dynamic range of the receiver was set so that the receiver can accommodate the useful signal $y(k)$ without distortion.

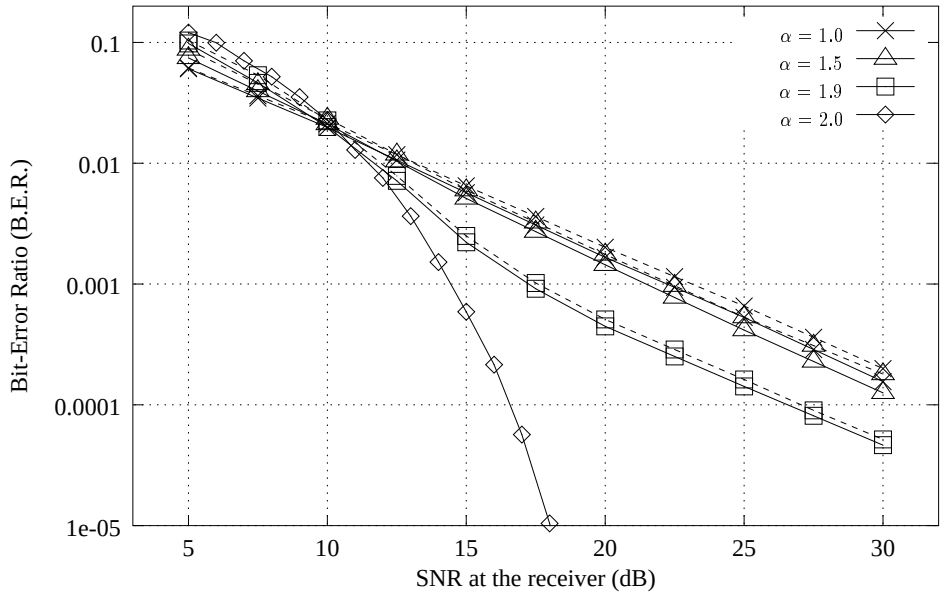


Figure 4.15: *Performance of the optimum (solid lines) and traditional (dashed lines) feed–forward Bayesian equalisers for channel $H_1(z)$ and a variety of values for α.*

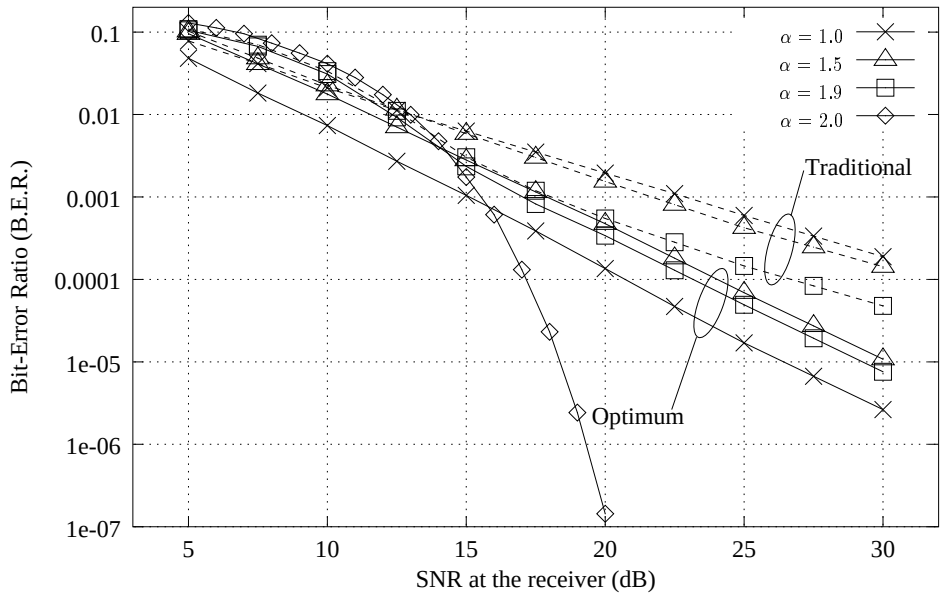


Figure 4.16: *Performance of the optimum (solid lines) and traditional (dashed lines) feed–forward Bayesian equalisers for channel $H_2(z)$ and a variety of values for α.*

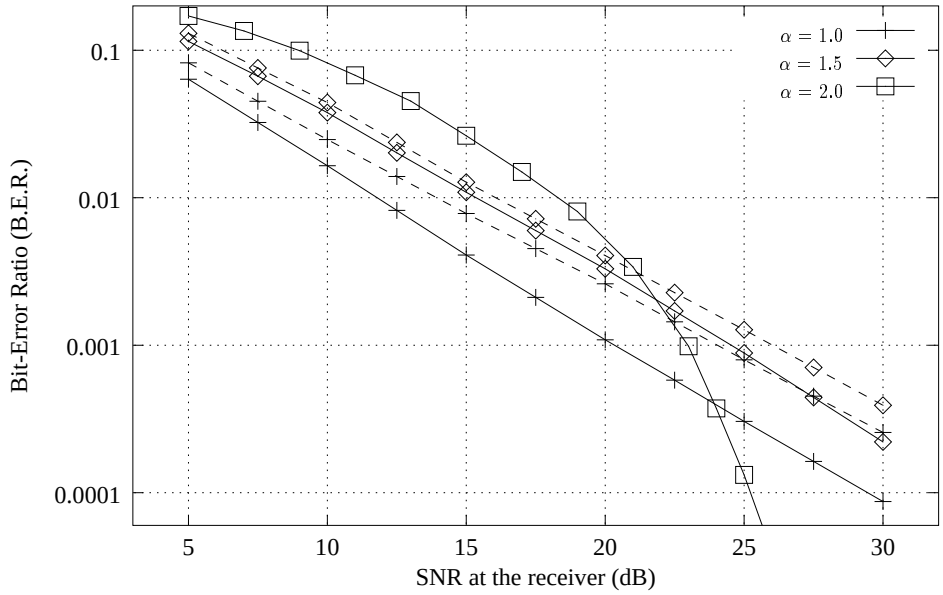


Figure 4.17: *Performance of the optimum (solid lines) and traditional (dashed lines) feed–forward Bayesian equalisers for a channel $H_3(z)$ and a variety of values for α.*

4.6.1 Feed–forward equalisers

For the first set of experiments the feed–forward Bayesian equaliser was simulated in varying noise environments. The equaliser order was $M = 2$ for all channels. Figure 4.15 shows the bit–error ratio (BER) performance of both optimum and traditional equalisers for channel $H_1(z)$ and a noise distribution ranging from Gaussian to Cauchy. The receiver dynamic range was $G = 3$ and the equaliser decision lag $d = 0$. The results suggest that for this channel the performance of the traditional equaliser is only marginally inferior to the optimum equaliser. That is because the decision boundaries for this scenario (which were depicted in fig. 4.2) change slightly with α. Specifically, the performance loss at 10^{-3} BER from the traditional equaliser is 0.32 dB for $\alpha = 1.9$, 0.89 dB for $\alpha = 1.5$ and 0.91 dB for $\alpha = 1$.

Figure 4.16 depicts the performance of the equalisers for the second channel $H_2(z)$ and the same range of values for α. The receiver dynamic range was again $G = 3$, but a decision lag of $d = 1$ was introduced in order to ensure the linear separability of the centers (see fig. 4.5 in page 52). The performance deviation of the traditional equaliser compared with the optimum one is significant in this case. The actual disadvantage is 1.03 dB for $\alpha = 1.9$, 3.78 dB for $\alpha = 1.5$ and 7.76 dB for $\alpha = 1$, at 10^{-3} BER. As fig. 4.5 indicates, the discrepancies between the Gaussian and non–Gaussian decision boundaries are substantial and result in an analogous performance degradation.

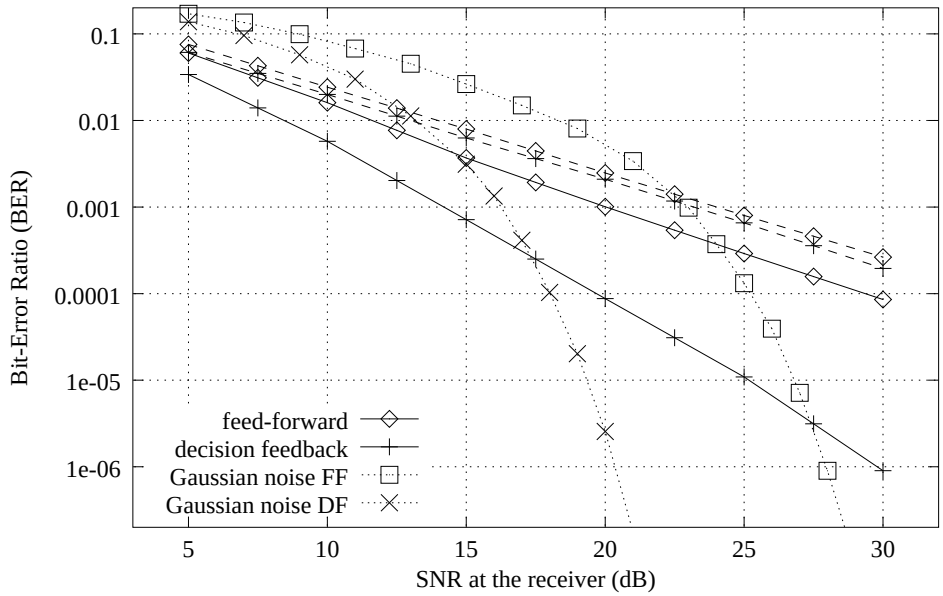


Figure 4.18: *Performance of the optimum (solid lines) and traditional (dashed lines) decision feedback Bayesian equaliser with the correct data fed back for* $\alpha = 1$ *(*$M = 2$*,* $D = 2$*,* $d = 1$*, and* $G = 4$*).*

For the last experiment concerning the feed–forward equalisers with channel $H_3(z)$, the dynamic range of the receiver was set to $G = 4$, and a decision lag $d = 1$ was adopted. The performance of the the optimally designed Bayesian equaliser (incorporating the α–stable noise pdf) was recorded, along with that of the traditional Bayesian equaliser (designed under the Gaussian assumption) for $\alpha = 2, 1.5$, and 1. The bit–error ratio for both equalisers is plotted in fig. 4.17. It can be clearly seen that the optimum MAP equaliser outperforms the traditional Bayesian equaliser when the noise deviates from the normal distribution. Namely, the performance benefit is 1.49 dB for $\alpha = 1.5$ and 3.71 dB for $\alpha = 1$ at a BER target of 10^{-3}.

4.6.2 Decision feedback equalisers

For this set of simulations, the performance of the optimum and traditional Bayesian decision feedback equaliser was recorded in a highly impulsive noise environment. Specifically, the characteristic exponent of the noise was set to $\alpha = 1$ (Cauchy distribution). The Gaussian noise case was also studied. The channel transfer function was $H_3(z)$, while the dynamic range of the receiver was set to $G = 4$, a value which allows the receiver to accommodate the useful signal without distortion. The order of the equalisers was $M = 2$ and they operated with a decision lag $d = 1$. The order of the feedback vector $\mathbf{x}_{\mathrm{D}}$ was chosen $D = 2$.

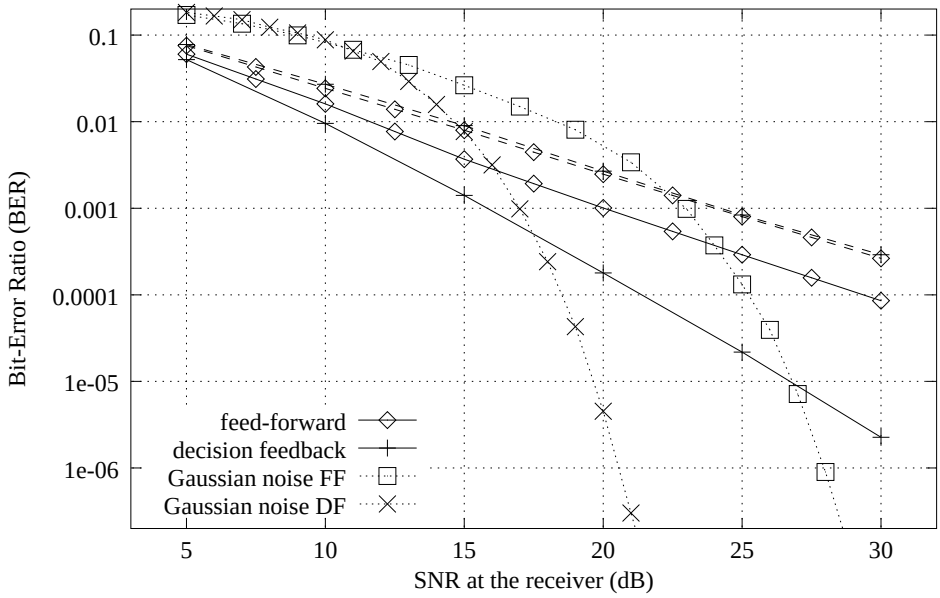


Figure 4.19: *Performance of the optimum (solid lines) and traditional (dashed lines) decision feedback Bayesian equaliser with the detected data fed back for* $\alpha = 1$ *(*$M = 2$*,* $D = 2$*,* $d = 1$*, and* $G = 4$*).*

Fig. 4.18 shows the performance of the equalisers in this highly impulsive α–stable noise environment. For comparison, the BER graphs of the decision feedback as well as the feed–forward MAP equalisers are plotted. In this experiment the correct transmitted data $x(k)$ were fed in the feedback vector $\mathbf{x}_{\mathrm{D}}$ (for the DF equalisers). The results show that at a BER target of 10^{-3}, the performance benefit of the optimum decision feedback equaliser is 8.99 dB, compared to the traditional Bayesian DFE. Consequently, the introduction of the decision feedback scheme resulted in a performance gain of 5.28 dB, compared to the feed–forward equaliser in fig. 4.17.

For the next experiment (fig. 4.19) the actual detected symbols $\hat{x}(k)$ from the DF equaliser were fed into the feedback vector. As expected, the performance of the DFE is now slightly inferior due to error propagation. However, the advantage of the optimum Bayesian DFE is still considerable, namely 8.24 dB at a 10^{-3} BER target. That is, the use of the actual detected data in the feedback vector results in a gain loss of only 0.75 dB at 10^{-3} BER.

It is interesting to notice that the actual shape of the BER graphs for non–Gaussian stable noise is inherently different from the traditional graphs in Gaussian noise. Indeed, the probability of a bit–error in a communication system is highly related to the probability of *exceedence* $P_{x>\tau}$ of the underlying noise distribution; that is, the probability that the random variable $\mathbf{x}$ *exceeds* (is greater than) a given level τ. As shown in equations (3.25) and (3.26), the probability of

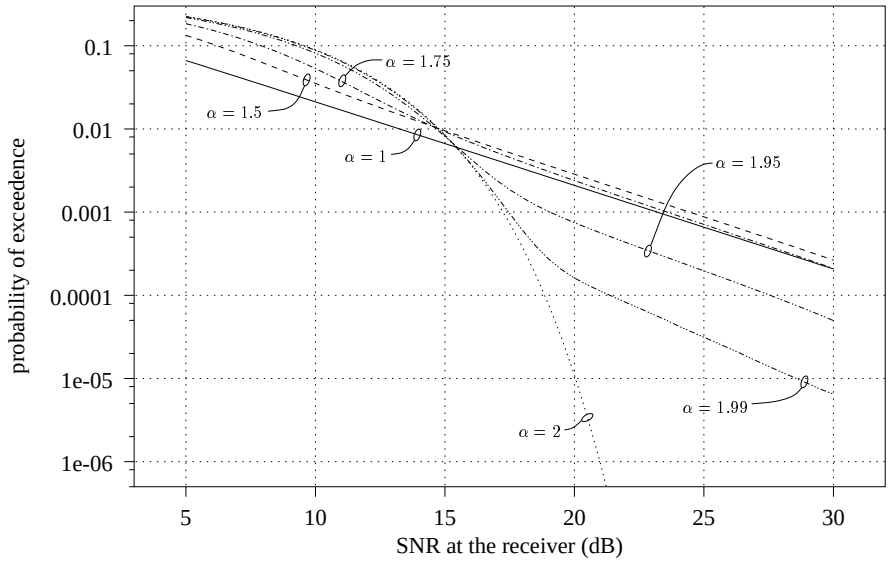


Figure 4.20: *Probability of exceedence of α–stable distribution for a variety of values for α ($\tau = .6$, $G = 4$).*

exceedence $P_{x>\tau}(\alpha, \gamma)$ for the Gaussian case ($\alpha = 2$) is

$$P_{x>\tau}(2, \gamma) = \frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\frac{\tau}{2\sqrt{\gamma}}\right) \tag{4.73}$$

and the Cauchy case ($\alpha = 1$)

$$P_{x>\tau}(1, \gamma) = \frac{1}{2} - \frac{\arctan\left(\frac{\tau}{\gamma}\right)}{\pi} \tag{4.74}$$

For all other values of α it is not possible to calculate this probability analytically, but we can still compute it experimentally.

Figure 4.20 depicts this probability as a function of the SNR at the receiver, for a variety of values of the characteristic exponent; the similarity with figures 4.15 to 4.19 is evident. Equation (4.71) was used here in order to map the noise dispersion γ to the corresponding values of SNR. In this mapping a receiver dynamic range $G = 4$ was assumed. The shape of the $P_{x>\tau}(\alpha, \gamma)$ graph is significantly different for Gaussian and non–Gaussian stable distributions, because of the property of eq. (3.24). This property (discussed in section 3.3, page 27) proves that while the Gaussian distribution has exponential tails, the stable laws have algebraic tails, i.e., tails which decay in an inverse power manner. Therefore, as in Mandelbrot's test, the plot of $\log P_{x>\tau}(\alpha, \gamma)$ against $\log \mathrm{SNR}$, as $\mathrm{SNR} \to \infty$, will be a straight line with slope $-\alpha$.

4.7 Conclusions

In this chapter, the optimum symbol–by–symbol equaliser for environments where the noise statistics exhibit stable characteristics was derived. This equaliser takes the form of a maximum *a posteriori* probabilities (MAP) detector, or Bayesian equaliser. And, in the same way as the stable law is a generalisation of the normal distribution, the optimum equaliser in stable noise environments turns out to be a generalised formulation of the traditional (Gaussian) Bayesian equaliser. The traditional Bayesian equaliser can be efficiently implemented with radial basis function (RBF) networks. However, it was demonstrated that when the noise is non–Gaussian α–stable, the basis functions are not radially symmetric, so the conventional RBF network, which evaluates the basis functions on scalar norms, can not be used. Nevertheless, a generalised structure, very similar to a RBF network can be utilised, which evaluates the multivariate noise distribution on vectors representing the distance of the input vector to the associated center. The feed–forward Bayesian equaliser is then extended introducing a decision feedback scheme, which exploits the information contained in the sequence of detected symbols, in order to reduce the intersymbol interference in the received signal.

The performance evaluation of the Bayesian equaliser in an α–stable noise environment with the use of conventional bit-error ratio against signal–to–noise ratio graphs is meaningless, due to the infinite variance of stable signals. However, as we suggested in Chapter 3, the infinite variance noise should be studied after being passed through the finite dynamic range front end of the receiver. Some limited analytical tools are consequently given in order to calculate the SNR *at the receiver*. This problem is generally intractable, because no closed form expressions exist for the α–stable density functions. Certain practical approximations, though, will be presented in Chapter 6.

Finally, a set of simulation experiments was conducted for the performance evaluation of the Bayesian equaliser in an α–stable noise environment. For these simulations the equalisers had perfect knowledge of the channel and noise characteristics, so the effects from inaccurate estimation of the equaliser parameters were not revealed. Nevertheless, the experimental results indicate that the proposed optimum Bayesian equaliser may offer significant performance benefits, with comparison to the Bayesian equaliser designed under the Gaussian assumption, in environments where the noise statistics deviate even slightly from the normal distribution.

Chapter 5
Training the equaliser

The optimum symbol–by–symbol equaliser derived in chapter 4 is fully defined by two sets of parameters: a) the vector centers $\mathbf{c}_i$ and their associated signs s_i ($0 \leq i < N_c$), and b) the parameters of the probability density function of the $S\alpha S$ noise, namely the characteristic exponent α and dispersion γ (see eq. (4.31), page 50 and fig. 4.8, page 58). The present chapter addresses the problem of determining these parameters for the Bayesian equaliser in a non–Gaussian α–stable noise environment.

For the estimation of the centers of a Bayesian equaliser two adaptive schemes have been developed in literature. The first approach [101] estimates the channel impulse response with a traditional linear adaptive algorithm and uses the resulting channel estimate to derive the equaliser centers. This approach, although computationally complex, needs a small training set to converge and is suitable for rapidly time–varying channels. The traditional channel estimation algorithms are based on second order statistics and are significantly weakened in non–Gaussian noise environments. In this scope, we present a family of recursive algorithms for robust channel estimation in a unified framework, highlighting the underlying relationships between them. The second scheme for estimating the equaliser centers identifies the vector centers directly using a clustering algorithm [76]. This scheme is computationally very simple and is immune from nonlinear channel distortion. However, it requires a large amount of training data and is only suitable for stationary or slowly time–varying channels. For this scheme we provide robust generalisations of the existing clustering algorithms, which have been developed in a Gaussian noise context and their performance is seriously degraded in an impulsive noise environment. Moreover, a novel clustering approach is proposed, based on robust location estimation for stable signals.

The estimation of the parameters of a $S\alpha S$ random variable is addressed in section 5.9. A variety of algorithms exist in the literature [58], but not all of them are suitable for a communications signal processing context. Section 5.9 presents three families of algorithms for the estimation of stable parameters, and provides experimental results for evaulating their performance.

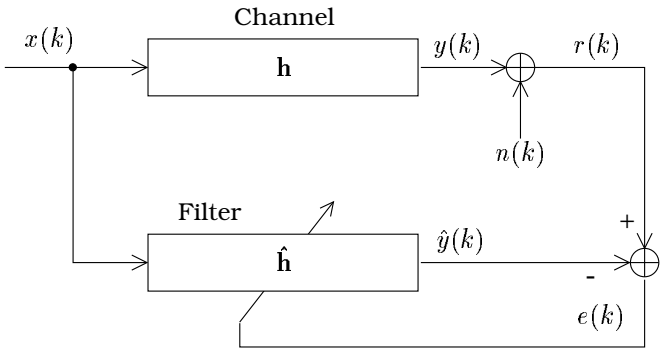


Figure 5.1: *Supervised channel identification model.*

5.1 System model

Consider the digital data sequence $\{x(k) = +1, -1\}$, consisting of independent and equiprobable binary symbols, which is passed through a noiseless linear channel with finite impulse response $\mathbf{h} = [h_0 \quad h_1 \quad \dots \quad h_{N-1}]^{\mathrm{T}}$. The output of the channel is

$$y(k) = \mathbf{h}^{\mathrm{T}}\mathbf{x}_{\mathrm{ch}}(k) \tag{5.1}$$

where

$$\mathbf{x}_{\mathrm{ch}}(k) = [x(k) \quad x(k-1) \quad \dots \quad x(k-N+1)]^{\mathrm{T}}$$

is the channel input vector. The observed sequence is corrupted by additive white noise $\{n(k)\}$ at the output of the channel

$$r(k) = y(k) + n(k) \tag{5.2}$$

The random noise sequence $\{n(k)\}$ is modelled as a symmetric α–stable process ($S\alpha S$).

The channel estimation problem in a supervised manner is depicted in fig. 5.1, and can be stated as follows: given the sequence of observations $\{r(k)\}$ and the respective channel input sequence $\{x(k)\}$, the task is to find an estimate

$$\hat{\mathbf{h}} = [\hat{h}_0 \quad \hat{h}_1 \quad \dots \quad \hat{h}_{N-1}]^{\mathrm{T}}$$

of the true channel tap weights vector $\mathbf{h}$. Here, we assume that the length of the channel impulse response is known, so that vector $\hat{\mathbf{h}}$ has the same size as vector $\mathbf{h}$. The channel estimation filter

performs the convolution of the channel input sequence $\{x(k)\}$ with the estimate $\hat{\mathbf{h}}$ to produce an estimate $\hat{y}(k)$ of the channel output, i.e.,

$$\hat{y}(k) = \hat{\mathbf{h}}^{\mathrm{T}}\mathbf{x}_{\mathrm{ch}}(k) = \sum_{i=0}^{N-1} \hat{h}_i x(k-i) \tag{5.3}$$

The estimation error is then

$$e(k) = r(k) - \hat{y}(k) = r(k) - \sum_{i=0}^{N-1} \hat{h}_i x(k-i) \tag{5.4}$$

and is used to drive the update algorithm of the channel estimation filter. According to section 4.5 (page 64) and fig. 4.11, we should use the limited received signal r_{L} in eq. (5.4), but in this chapter we shall ignore the limiter at the front end of the receiver.

There is a variety of adaptive algorithms for updating the channel impulse response estimate $\hat{\mathbf{h}}$ [35]. Two are the most dominant families among these: a) the stochastic gradient algorithms, and b) the least squares algorithms. Both families produce an estimate for the channel impulse response which minimises an appropriate cost function of the estimation error.

5.2 Stochastic least mean p–norm (LMP)

The optimisation criterion in Gaussian noise environments, for linear regression problems, is usually the minimisation of a quadratic function of the estimation error. One of the most widely used is the *minimum mean squared error* (MMSE) criterion, which adopts a cost function of the form

$$J_{\mathrm{MMSE}} = \boldsymbol{E}\left\{e^2(k)\right\} \tag{5.5}$$

The minimisation of the MMSE cost function leads to the *Wiener* equation [35] for the optimum vector of tap weights

$$\hat{\mathbf{h}}_{\mathrm{opt}} = \boldsymbol{\phi}_{xx}^{-1}\boldsymbol{\phi}_{xr} \tag{5.6}$$

where $\boldsymbol{\phi}_{xx}$ is the autocorrelation matrix of the input vector $\mathbf{x}_{\mathrm{ch}}(k)$, and $\boldsymbol{\phi}_{xr}$ is the cross–correlation vector between the input vector and the desired response $r(k)$. A very simple and

widely used iterative algorithm which, in the mean, provides the Wiener estimate is the least mean squares (LMS) algorithm [35, 102]. It can be summarised as:

Algorithm 5.1 (LMS)
Initialisation: $\hat{\mathbf{h}}(0) = \mathbf{0}$
Basic recursion: $\forall\ k = 1, 2, \ldots$

1. *a priori* estimation error

$$\xi(k) = r(k) - \hat{\mathbf{h}}^{\mathrm{T}}(k-1)\mathbf{x}_{\mathrm{ch}}(k)$$

2. tap weight adaptation

$$\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k-1) + \mu\,\xi(k)\,\mathbf{x}_{\mathrm{ch}}(k) \tag{5.7}$$

3. repeat from step 1

□

where μ is the *step–size parameter*, which controls the adaptation rate. The *a priori* estimation error $\xi(k)$ is defined as

$$\xi(k) = r(k) - \hat{\mathbf{h}}^{\mathrm{T}}(k-1)\mathbf{x}_{\mathrm{ch}}(k) \tag{5.8}$$

where the inner product $\hat{\mathbf{h}}^{\mathrm{T}}(k-1)\mathbf{x}_{\mathrm{ch}}(k)$ represents an estimate of the desired response $r(k)$, based on the *old* estimate of the tap weight $\hat{\mathbf{h}}^{\mathrm{T}}(k-1)$ vector, that was made at time $k-1$.

As fig. 5.2 shows, the LMS can efficiently estimate the taps of channel $H(z) = 1 + 0.5z^{-1}$ in a Gaussian noise environment. Here, the step–size parameter was $\mu = 0.004$, while the noise dispersion $\gamma = 0.08$ which corresponds to a variance $\sigma^2 = 0.16$. The performance of LMS, however, degrades seriously when the statistics of the noise deviate from the Gaussian assumption. For example, fig. 5.3 depicts the channel tap estimates produced by LMS for the same constellation as fig. 5.2, but with the characteristic exponent of the noise set to $\alpha = 1$. It is evident that the algorithm diverges due to the frequent occurrence of large noise samples.

This behaviour of LMS is not surprising. Recall from chapter 3, that for a non–Gaussian stable distribution with characteristic exponent α, only moments of order less than α are finite. In

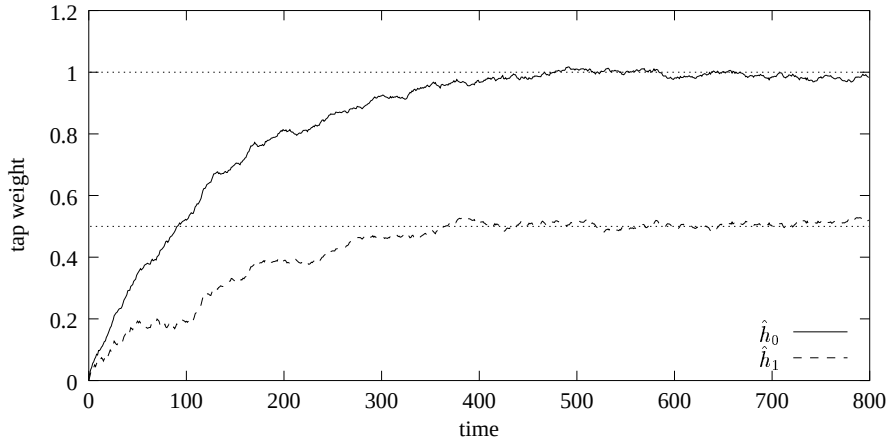


Figure 5.2: *A single LMS run in Gaussian noise for channel* $H(z) = 1 + 0.5z^{-1}$ *(*$\alpha = 2, \gamma = 0.08$*).*

particular, the variance (i.e. the second order moment) of a stable distribution with $\alpha < 2$ does not exist, making the use of variance or correlation as measures of dispersion (as in eq. (5.5)) meaningless.

As shown in [9], the *dispersion* of a stable random variable plays an analogous role of the variance. For example, the larger the dispersion of a stable distribution, the more it spreads around its location parameter. Hence the *minimum dispersion* criterion becomes a natural and mathematically meaningful choice as a measure of optimality in stable signal processing. By minimising the error dispersion, we minimise the average magnitude of the estimation error. Furthermore, it has been shown that minimising the dispersion is also equivalent to minimising the probability of large estimation errors [70].

The analytical extraction, though, of a solution to a minimum dispersion problem is, in many ways, mathematically intractable. However, adaptive solutions of linear estimation problems for stable processes are practically easy to implement, because they do not require closed form expressions. The dispersion of the estimation error is usually a convex function of the parameters. So, numerical methods, such as stochastic gradient methods, may be used to find the parameters by minimising the dispersion of the error function.

Specifically, we would like to find a vector

$$\hat{\mathbf{h}} = [\hat{h}_0 \quad \hat{h}_1 \quad \hat{h}_2 \quad \dots \quad \hat{h}_{N-1}]^{\mathrm{T}}$$

such that the dispersion of the error $e(k)$ is minimised. The minimum dispersion (MD) cost

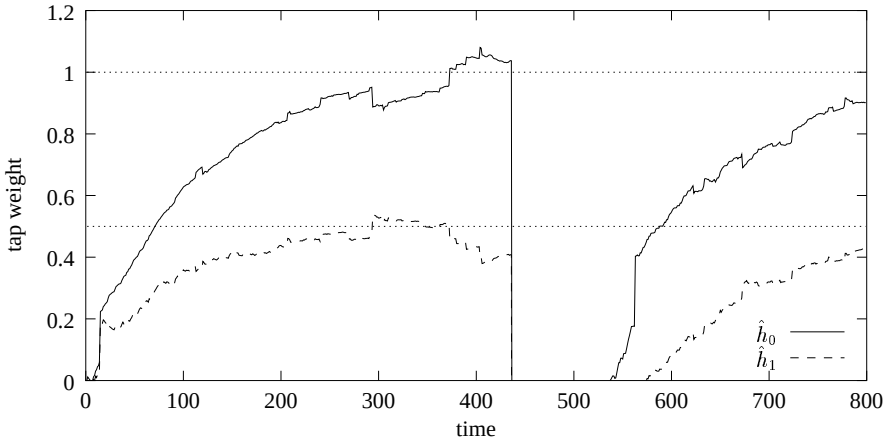


Figure 5.3: *A single LMS run in* SαS *noise for channel* $H(z) = 1 + 0.5z^{-1}$ *(*$\alpha = 1, \gamma = 0.08$*).*

function is thus given by

$$J_{\text{MD}} = \left\| r(k) - \sum_{i=0}^{N-1} \hat{h}_i x(k-i) \right\|_{\alpha} \tag{5.9}$$

where $\| \cdot \|_{\alpha}$ is the α–stable norm, defined in eq. (3.48). This cost function turns out to be analytically intractable, but an equivalent form can be used. According to Theorem 3.8, the norm of a *S*α*S* random variable is proportional to its p–th order moment for any $0 < p < \alpha$. So, an equivalent minimum mean p–norm cost function of the error (MMPE) is given by

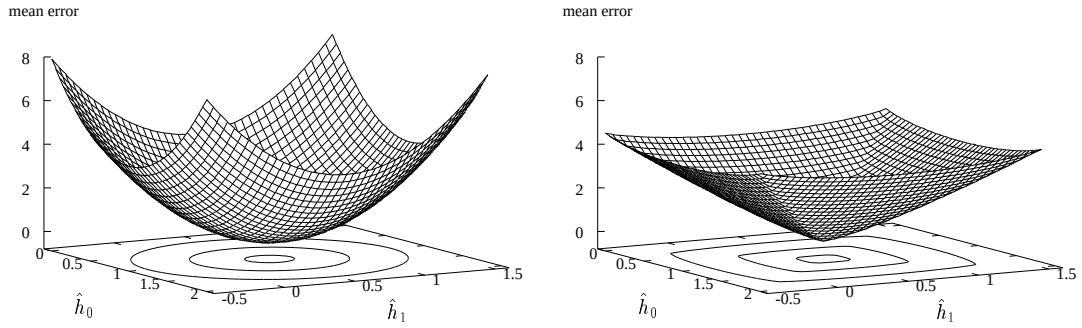
$$J_{\text{MMPE}} = \boldsymbol{E} \left\{ \left| r(k) - \sum_{i=0}^{N-1} \hat{h}_i x(k-i) \right|^p \right\} \tag{5.10}$$

where $0 < p < \alpha$. There is no closed form solution for the set of coefficients that minimise the cost function J_{MMPE} in eq. (5.10). But, J_{MMPE} is convex and so we may use a stochastic gradient method to solve for the coefficients in a similar way as the least mean square (LMS) [35] algorithm does. For example, fig. 5.4 depicts the error hyper–surface defined by J_{MMPE} for channel $H(z) = 1 + 0.5z^{-1}$ and two values of the parameter p. In [9] the authors propose the following generalisation of LMS, called least mean p–norm (LMP) algorithm:

Algorithm 5.2 (LMP) Fix p so that $1 \leq p < \alpha$.

Initialisation: $\hat{\mathbf{h}}(0) = \mathbf{0}$

Basic recursion: $\forall\ k = 1, 2, \ldots$



(a) $p = 2$ **(b)** $p = 1.2$

Figure 5.4: *Error surface of the least p–norm criterion for two values of p. The channel is* $H(z) = 1 + 0.5z^{-1}$.

1. *a priori* estimation error

$$\xi(k) = r(k) - \hat{\mathbf{h}}^{\mathrm{T}}(k-1)\mathbf{x}_{\mathrm{ch}}(k)$$

2. tap weight adaptation

$$\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k-1) + \mu \left\langle \xi(k) \right\rangle^{p-1} \mathbf{x}_{\mathrm{ch}}(k) \tag{5.11}$$

3. repeat from step 1

□

where $\mu > 0$ is the step–size parameter and the function $\langle \cdot \rangle$ is defined as

$$\langle x \rangle^{p-1} = \mathrm{sgn}(x)\, |x|^p \tag{5.12}$$

When $p = 1$, the above algorithm is called the LMAD (least mean absolute deviation) algorithm.

As an example, we consider the same configuration as fig. 5.3, but we use LMP in place of LMS. As fig. 5.5 shows, the channel taps estimates produced by LMP are much more reliable than those of LMS in an impulsive noise environment, while the convergence speed is comparable with that of LMS with Gaussian noise. Further experimental results conducted in [9] show clearly that the LMP and LMAD algorithms outperform the traditional LMS algorithm when

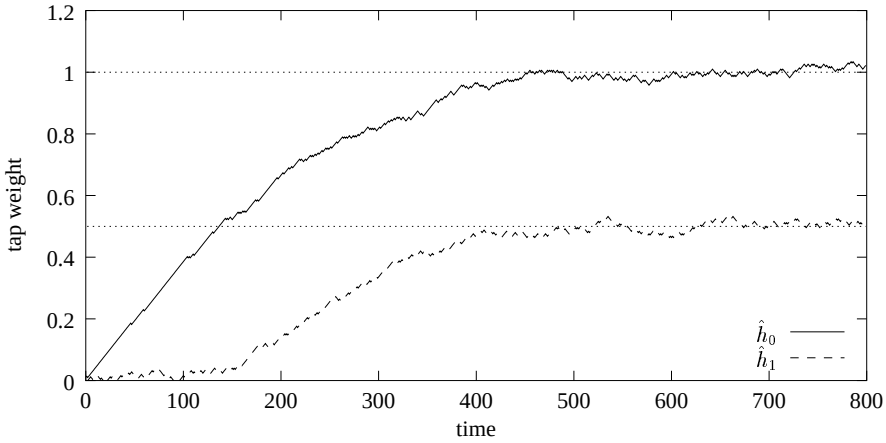


Figure 5.5: *A single run for LMP in* SαS *noise (*$\alpha = 1.0, \gamma = 0.08$*).*

the interfering noise is even slightly far from Gaussian. In particular, in a Gaussian environment ($\alpha = 2$), the LMAD is slower than the LMS. This situation changes dramatically when α is less than 2. As α decreases, the LMS becomes slower and slower to converge. In fact, when α is close to 1, the LMS hardly converges without its step size parameter approximately equal to zero. On the other hand, the LMAD and LMP maintain fairly constant rate of convergence over the whole range of α. The LMAD converges faster and faster relative to the LMS as α becomes smaller and smaller while maintaining the same misadjustment as that of the LMS. On the other hand, LMAD is simpler to implement than the LMS, since the tap weight update equation is simpler. The normalisation of the aforementioned algorithms has been studied in [103]. The new proposed algorithms are called NLMP (normalised least mean p–norm) and NLMAD (normalised least mean absolute deviation) algorithms. The normalised versions exhibit significantly superior convergence properties, compared to the original LMP and LMAD algorithms.

The stochastic gradient algorithms share the very desirable characteristic of simplicity. The need, however, for faster converging algorithms often calls for least squares (LS) signal processing. The effort to provide robust versions of LS algorithms for non–Gaussian noise environments has received attention in the literature [7, 104–106]. In the following sections (5.3 to 5.5) we present a family of robust LS type algorithms for adaptive channel identification under a unified framework [107]. A modified version of LMP is also presented in section 5.6, under the light of the findings in the following sections. The main characteristics of these algorithms is their robustness in non–Gaussian environments, and a complexity which is comparable to the complexity of the traditional linear LS algorithms.

5.3 The least squares approach

The *least squares* criterion has been adopted in Gaussian signal processing, instead of the statistical MMSE criterion, to derive faster converging adaptive filtering algorithms. Least squares actually deal directly with the data sequences and obtain estimates of correlations from the data itself. More specifically, the expectation operator in eq. (5.5) is replaced by a time summation. Thus, the cost function a time k, is the weighted sum of squared errors

$$J_{\mathrm{LS}} = \sum_{i=0}^{k} \lambda^{k-i}\, e^2(i) \tag{5.13}$$

where λ ($0 < \lambda \leq 1$) is an exponential weighting factor. Its purpose is to weight the most recent data more heavily and thus allow the filter coefficients to adapt to time–varying channel scenarios. This exponential weighting of the error samples corresponds to an equivalent window of size $W = 1/(1-\lambda)$ [35]. Just as the LS cost function can be obtained from the MMSE cost function (eq. (5.5)), $\hat{\mathbf{h}}(k)$ can be obtained from eq. (5.6) by replacing expectation (for the correlation matrices) with time summation. Indeed, in order to minimise eq. (5.13) we differentiate J_{LS} with respect to each of the filter coefficients $\hat{h}_j$

$$\begin{aligned} \frac{\partial J}{\partial \hat{h}_j} &= \sum_{i=0}^{k} \lambda^{k-i} \frac{\partial e^2(i)}{\partial \hat{h}_j} \\ &= -2 \sum_{i=0}^{k} \lambda^{k-i} e(i)\, x(i-j), \quad j = 0,\, 1,\, \ldots\, N-1 \end{aligned} \tag{5.14}$$

Consequently, minimisation of eq. (5.13) implies

$$\sum_{i=0}^{k} \lambda^{k-i}\, e(i)\, x(i-j) = 0, \quad \forall\, j = 0,\, 1,\, \ldots\, N-1 \tag{5.15}$$

This is a set of N linear equations with N unknowns. Replacing the estimation error from eq. (5.4) in eq. (5.15), and exchanging the summation order we obtain

$$\sum_{l=0}^{N-1} \hat{h}_l \sum_{i=0}^{k} \lambda^{k-i}\, x(i-l)\, x(i-j) = \sum_{i=0}^{k} \lambda^{k-i}\, x(i-j)\, r(i), \quad \forall\, j = 0, 1, \ldots, N-1 \tag{5.16}$$

It is, however, more convenient to express the above set of equations in matrix form

$$\left[\sum_{i=0}^{k} \lambda^{k-i}\, \mathbf{x}_{\text{ch}}(i)\, \mathbf{x}_{\text{ch}}^{\text{T}}(i)\right]\, \hat{\mathbf{h}}(k) = \left[\sum_{i=0}^{k} \lambda^{k-i}\, \mathbf{x}_{\text{ch}}(i)\, r(i)\right] \tag{5.17}$$

We now define the *sample* autocorrelation matrix of the data $\mathbf{R}_{xx}$ and cross–correlation vector $\mathbf{r}_{xr}$ as

$$\mathbf{R}_{xx}(k) = \sum_{i=0}^{k} \lambda^{k-i}\, \mathbf{x}_{\text{ch}}(i)\, \mathbf{x}_{\text{ch}}^{\text{T}}(i) \tag{5.18}$$

$$\mathbf{r}_{xr}(k) = \sum_{i=0}^{k} \lambda^{k-i}\, \mathbf{x}_{\text{ch}}(i)\, r(i) \tag{5.19}$$

or in recursive form

$$\mathbf{R}_{xx}(k) = \lambda\, \mathbf{R}_{xx}(k-1) + \mathbf{x}_{\text{ch}}(k)\, \mathbf{x}_{\text{ch}}^{\text{T}}(k) \tag{5.20}$$

$$\mathbf{r}_{xr}(k) = \lambda\, \mathbf{r}_{xr}(k-1) + \mathbf{x}_{\text{ch}}(k)\, r(k) \tag{5.21}$$

Therefore, eq. (5.17) can be written in a more compact way as follows

$$\mathbf{R}_{xx}(k)\, \hat{\mathbf{h}}(k) = \mathbf{r}_{xr}(k) \tag{5.22}$$

To compute the least squares estimate $\hat{\mathbf{h}}(k)$ for the tap–weight vector in accordance with eq. (5.22), we have to determine the inverse of the sample autocorrelation matrix $\mathbf{R}_{xx}(k)$. In practice, we usually try to avoid performing such an operation as it can be very time consuming, particularly if the number of tap weights N is high. Also, we would like to be able to update the least squares estimate recursively, as new data samples arrive. We can realise both of these objectives by using a basic result in matrix algebra known as the *matrix inversion lemma* [35, 102].

Lemma 5.1 (The Matrix Inversion Lemma) *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $(N \times N)$ *positive–definite matrices related by*

$$\mathbf{A} = \mathbf{B}^{-1} + \mathbf{C}\mathbf{D}^{-1}\mathbf{C}^{\text{T}} \tag{5.23}$$

where $\mathbf{D}$ *is another* $(L \times N)$ *positive–definite matrix, and* $\mathbf{C}$ *is a* $(N \times L)$ *matrix. Then, we*

may express the inverse of the matrix $\mathbf{A}$ *as follows:*

$$\mathbf{A}^{-1} = \mathbf{B} - \mathbf{BC}(\mathbf{D} + \mathbf{C}^T\mathbf{BC})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{B} \tag{5.24}$$

□

Applying now this lemma to the recursive equation eq. (5.20) we derive the recursive least squares (RLS) algorithm [1] which can be summarised as follows:

Algorithm 5.3 (RLS)

Initialisation: $\mathbf{P}(0) = \delta^{-1}\mathbf{I}, \quad \hat{\mathbf{h}}(0) = \mathbf{0}$

Basic recursion: $\forall\ k = 1, 2, \ldots$

1. *a priori* estimation error

$$\xi(k) = r(k) - \hat{\mathbf{h}}^{\mathrm{T}}(k-1)\,\mathbf{x}_{\mathrm{ch}}(k)$$

2. Kalman gain

$$\mathbf{K}(k) = \frac{\lambda^{-1}\,\mathbf{P}(k-1)\,\mathbf{x}_{\mathrm{ch}}(k)}{1 + \lambda^{-1}\,\mathbf{x}_{\mathrm{ch}}^{\mathrm{T}}\,\mathbf{P}(k-1)\,\mathbf{x}_{\mathrm{ch}}(k)}$$

3. tap weights update

$$\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k-1) + \xi(k)\,\mathbf{K}(k)$$

4. inverse autocorrelation matrix

$$\mathbf{P}(k) = \lambda^{-1}\mathbf{P}(k-1) - \lambda^{-1}\,\mathbf{K}(k)\,\mathbf{x}_{\mathrm{ch}}^{\mathrm{T}}\,\mathbf{P}(k-1)$$

5. repeat from step 1

□

The role of the least squares criterion under the stable assumption, however, is questionable. In the first place, the stable distribution is best used to model signals and noise that exhibit

[1] The full algebra for the derivation of RLS can be found in [35, 108]

impulsive nature. This type of signals tends to produce outliers. Although the least squares criterion is adequate under the Gaussian assumption and usually leads to analytically tractable results, it is not appropriate for an impulsive non–Gaussian environment, largely due to its non–robustness against outliers [7, 67, 68, 104, 105]. On the other hand, the use of second order statistics in eq. (5.13) is meaningless in the infinite variance environment of α–stable noise. Indeed, experimental simulations indicate that the behaviour of RLS in a non–Gaussian stable noise environment is similar to that of LMS, as shown in fig. 5.3.

5.4 Order selective recursive least squares (OSRLS)

Based on the aforementioned arguments, one can devise a heuristic method to deal with the outliers of non–Gaussian distributions. A traditional LS algorithm (e.g. RLS [35]) is still used, but the channel estimate adaptation is inhibited when the received signal $r(k)$ is corrupted by a noise sample which can be characterised as an *outlier*. In order to identify these *outliers*, order statistics are used. Consider a sorted vector containing the magnitude of the last Θ estimation error samples $\xi(k)$. If the current error sample lies among the top η largest samples in this vector, the current observation is characterised as an *outlier*. For example, an *order selective RLS* (OSRLS) algorithm can be expressed as follows:

Algorithm 5.4 (OSRLS)

Initialisation: $\mathbf{P}(0) = \delta^{-1}\mathbf{I}, \quad \hat{\mathbf{h}}(0) = \mathbf{0}$

Basic recursion: $\forall\ k = 1, 2, \ldots$

1. *a priori* estimation error

$$\xi(k) = r(k) - \hat{\mathbf{h}}^{\mathrm{T}}(k-1)\,\mathbf{x}_{\mathrm{ch}}(k)$$

2. sorted error magnitude vector

$$\boldsymbol{\xi}_{\mathrm{s}}(k) \triangleq \mathrm{sort}([|\xi(k)| \quad |\xi(k-1)| \quad \ldots \quad |\xi(k-\Theta+1)|])$$

3. find the rank of $|\xi(k)|$ in $\boldsymbol{\xi}_{\mathrm{s}}(k)$

$$\mathcal{R}(k) = \mathrm{rank}_{\boldsymbol{\xi}_{\mathrm{s}}(k)}(|\xi(k)|)$$

4. if $\mathcal{R}(k) \leq \Theta - \eta$, then perform the RLS recursion

5. repeat from step 1

□

As the experimental results in section 5.7 suggest, OSRLS can achieve good performance in highly impulsive environments. Its main disadvantage, though, is that there is no way of determining the optimal values for the parameters η and Θ.

5.5 Recursive weighted least squares (RWLS)

The class of *M*–estimators is a robust version of the LS estimate, proposed by Huber [106]. Instead of minimising the sum of squared error, a less rapidly increasing function $\rho(\cdot)$ of the error is used. The corresponding cost function is then

$$\mathcal{J}_M = \sum_{i=0}^{k} \lambda^{k-i} \rho(e(i)) \tag{5.25}$$

The parameter λ ($0 < \lambda \leq 1$) is the exponential weighting factor. In order to minimise eq. (5.25) we differentiate $\mathcal{J}_M$ with respect to each of the filter coefficients $\hat{h}_j$

$$\begin{aligned} \frac{\partial \mathcal{J}_M}{\partial \hat{h}_j} &= \sum_{i=0}^{k} \lambda^{k-i} \frac{\partial \rho(e(i))}{\partial \hat{h}_j} \\ &= -\sum_{i=0}^{k} \lambda^{k-i} \psi(e(i))\, x(i-j) \quad , \quad j = 0\,, 1\,, \ldots\, N-1 \end{aligned} \tag{5.26}$$

where $\psi = \rho'$ is the derivative of ρ. Hence, the minimisation of eq. (5.25) implies

$$\sum_{i=0}^{k} \lambda^{k-i} \psi(e(i)) x(i-j) = 0 \quad , \quad \forall\, j = 0,\, 1,\, \ldots\, N-1 \tag{5.27}$$

Now, let

$$\phi(x) = \frac{\psi(x)}{x}, \quad \text{and} \quad v(i) = \phi(e(i)) \tag{5.28}$$

Equation (5.27) can be rewritten as follows:

$$\sum_{i=0}^{k} \lambda^{k-i} v(i) e(i) x(i-j) = 0, \quad \forall\, j = 0, 1, \ldots N-1 \tag{5.29}$$

Equation (5.29) may be viewed as a *soft–selective* version (compared to the *hard* selection of OSRLS) of the traditional LS techniques. The summation terms are not all incorporated linearly in the summation, but the smaller ones are weighted more heavily, depending on the specific penalty function ρ. The actual weighting is controlled by the *weighting function* ϕ, defined in eq. (5.28). On the other hand, eq. (5.29) can be rewritten as a generalisation of the traditional LS equations (eq. (5.15)), as follows:

$$\sum_{i=0}^{k} \lambda^{k-i} e_{\mathrm{g}}(i) x_{\mathrm{g}}(i-j) = 0, \quad \forall\, j = 0, 1, \ldots N-1 \tag{5.30}$$

where

$$\begin{aligned} x_{\mathrm{g}}(i-j) &= v^{\frac{1}{2}}(i) x(i-j) \\ r_{\mathrm{g}}(i) &= v^{\frac{1}{2}}(i) r(i) \\ e_{\mathrm{g}}(i) &= v^{\frac{1}{2}}(i) e(i) \end{aligned} \tag{5.31}$$

Equation (5.30) is essentially a *generalised orthogonality criterion*. This generalisation implies (as in [109]) that all quadratic adaptive filter algorithms can be made robust if the vector $\mathbf{x}_{\mathrm{ch}}(i)$ and the desired response $r(i)$ at time i are modified according to eq. (5.31). In fact, LMP can be derived from LMS using this generalisation.

After some rearrangements and exchange in summation order in eq. (5.29) we obtain

$$\begin{aligned} \sum_{l=0}^{N-1} \hat{h}_l \sum_{i=0}^{k} \lambda^{k-i}\, v(i)\, x(i-l)\, x(i-j) &= \sum_{i=0}^{k} \lambda^{k-i}\, v(i)\, x(i-j)\, r(i) \\ & \forall\, j = 0,\, 1,\, \ldots\, N-1 \end{aligned} \tag{5.32}$$

It is, however, more convenient to express the above set of equations in matrix form

$$\left[\sum_{i=0}^{k} \lambda^{k-i} v(i) \mathbf{x}_{\mathrm{ch}}(i) \mathbf{x}_{\mathrm{ch}}^{\mathrm{T}}(i) \right] \hat{\mathbf{h}}(k) = \left[\sum_{i=0}^{k} \lambda^{k-i} v(i) \mathbf{x}_{\mathrm{ch}}(i) r(i) \right] \tag{5.33}$$

Equations (5.32) or (5.33) are not a set of linear equations in the unknown parameters $\hat{\mathbf{h}}$. We have, however, reformulated these equations to appear linear in $\hat{\mathbf{h}}$ by hiding the nonlinearity in the data weighting term

$$v(i) = \phi(e(i)) \tag{5.34}$$

But the sequence $v(i)$ assumes knowledge of the optimal (in the sense of eq. (5.25)) weight vector $\hat{\mathbf{h}}_{\text{opt}}(k)$ at time k to generate the error sequence $e(i)$ associated with that tap weight vector, since

$$e(i) = y(i) - \hat{\mathbf{h}}_{\text{opt}}^{\text{T}}(k)\mathbf{x}_{\text{ch}}(i) \tag{5.35}$$

However, we can assume that the tap–weight estimate $\hat{\mathbf{h}}$ changes slowly with time. Therefore, as in [104, 110], for the recursive approximation of $\hat{\mathbf{h}}$, the instantaneous *a priori* estimation error $\xi(i) = y(i) - \hat{\mathbf{h}}^{\text{T}}(i-1)\mathbf{x}_{\text{ch}}(i)$ may be used to approximate $e(i)$. We can, therefore, generate the sequence

$$w(i) = \phi(\xi(i)) \tag{5.36}$$

in order to approximate $v(i)$. We now define the sample *weighted autocorrelation matrix* of the data $\mathbf{Q}_{xx}$ and the *weighted cross–correlation vector* $\mathbf{q}_{xr}$ as [72, 105, 106]

$$\mathbf{Q}_{xx}(k) = \sum_{i=0}^{k} \lambda^{k-i}\, w(i)\, \mathbf{x}_{\text{ch}}(i)\, \mathbf{x}_{\text{ch}}^{\text{T}}(i) \tag{5.37}$$

$$\mathbf{q}_{xr}(k) = \sum_{i=0}^{k} \lambda^{k-i}\, w(i)\, \mathbf{x}_{\text{ch}}(i)\, r(i) \tag{5.38}$$

or, in recursive form

$$\mathbf{Q}_{xx}(k) = \lambda\mathbf{Q}_{xx}(k-1) + w(k)\mathbf{x}_{\text{ch}}(k)\mathbf{x}_{\text{ch}}^{\text{T}}(k) \tag{5.39}$$

$$\mathbf{q}_{xr}(k) = \lambda\mathbf{q}_{xr}(k-1) + w(k)\mathbf{x}_{\text{ch}}(k)r(k) \tag{5.40}$$

Using these definitions, we may use the following approximation to eq. (5.33)

$$\mathbf{Q}_{xx}(k)\hat{\mathbf{h}}(k) = \mathbf{q}_{xr}(k) \tag{5.41}$$

To compute the M–estimate $\hat{\mathbf{h}}(k)$ for the tap–weight vector in accordance with eq. (5.41) , we have to determine the inverse of the *weighted* autocorrelation matrix $\mathbf{Q}_{xx}(k)$. As in section 5.3 we will proceed by applying the *matrix inversion lemma* (eq. (5.24)) in eq. (5.39). We first make the following identifications:

$$\begin{aligned} \mathbf{A} &= \mathbf{Q}_{xx}(k) \\ \mathbf{B} &= \lambda^{-1}\mathbf{Q}_{xx}^{-1}(k-1) \\ \mathbf{C} &= \mathbf{x}_{\text{ch}}(k) \\ \mathbf{D} &= w(k)^{-1} \end{aligned}$$

Then, substituting these definitions in eq. (5.24), we obtain the following recursive formula for the inverse of the weighted autocorrelation matrix

$$\mathbf{Q}_{xx}^{-1}(k) = \lambda^{-1}\mathbf{Q}_{xx}^{-1}(k-1) - \frac{\lambda^{-2}\mathbf{Q}_{xx}^{-1}(k-1)\mathbf{x}_{\text{ch}}(k)\mathbf{x}_{\text{ch}}^{\text{T}}(k)\mathbf{Q}_{xx}^{-1}(k-1)}{w(k)^{-1} + \lambda^{-1}\mathbf{x}_{\text{ch}}^{\text{T}}(k)\mathbf{Q}_{xx}^{-1}(k-1)\mathbf{x}_{\text{ch}}(k)} \tag{5.42}$$

For convenience, let

$$\mathbf{P}(k) = \mathbf{Q}_{xx}^{-1}(k) \tag{5.43}$$

and

$$\mathbf{K}(k) = \frac{\lambda^{-1}\mathbf{P}(k-1)\mathbf{x}_{\text{ch}}(k)}{w(k)^{-1} + \lambda^{-1}\mathbf{x}_{\text{ch}}^{\text{T}}(k)\mathbf{P}(k-1)\mathbf{x}_{\text{ch}}(k)} \tag{5.44}$$

Using these definitions, we may rewrite eq. (5.42) as follows

$$\mathbf{P}(k) = \lambda^{-1}\mathbf{P}(k-1) - \lambda^{-1}\mathbf{K}(k)\mathbf{x}_{\text{ch}}^{\text{T}}(k)\mathbf{P}(k-1) \tag{5.45}$$

The $(N \times N)$ matrix $\mathbf{P}(k)$ will be referred to as the *inverse weighted correlation matrix*. The N–vector $\mathbf{K}(k)$ will be referred to as the *gain vector*. By rearranging eq. (5.44), we have

$$\mathbf{K}(k) = w(k)\ \left(\lambda^{-1}\mathbf{P}(k-1) - \lambda^{-1}\mathbf{K}(k)\mathbf{x}_{\text{ch}}^{\text{T}}(k)\mathbf{P}(k-1)\right)\ \mathbf{x}_{\text{ch}}(k) \tag{5.46}$$

which also implies

$$\mathbf{K}(k) = w(k)\ \mathbf{P}(k)\ \mathbf{x}_{\text{ch}}(k) \tag{5.47}$$

Next, we will develop a recursive equation for updating the M–estimate $\hat{\mathbf{h}}(k)$ for the tap–weight vector. To do this, we use eqs. (5.40), (5.41), and (5.43) to express the estimate $\hat{\mathbf{h}}(k)$ at time k as follows:

$$\begin{aligned}\hat{\mathbf{h}}(k) &= \mathbf{Q}_{xx}^{-1}(k)\,\mathbf{q}_{xr}(k) \\ &= \mathbf{P}(k)\,\mathbf{q}_{xr}(k) \\ &= \lambda\,\mathbf{P}(k)\,\mathbf{q}_{xr}(k-1) + w(k)\,\mathbf{P}(k)\,\mathbf{x}_{\text{ch}}(k)\,r(k)\end{aligned} \tag{5.48}$$

Substituting eq. (5.45) for $\mathbf{P}(k)$ in the first term only in the right–hand side of eq. (5.48) we have

$$\begin{aligned}\hat{\mathbf{h}}(k) &= \mathbf{P}(k-1)\,\mathbf{q}_{xr}(k-1) - \mathbf{K}(k)\,\mathbf{x}_{\text{ch}}^{\text{T}}(k)\,\mathbf{P}(k-1)\,\mathbf{q}_{xr}(k-1) \\ &\quad + w(k)\,\mathbf{P}(k)\,\mathbf{x}_{\text{ch}}(k)\,r(k) \\ &= \mathbf{Q}_{xx}^{-1}(k-1)\,\mathbf{q}_{xr}(k-1) - \mathbf{K}(k)\,\mathbf{x}_{\text{ch}}^{\text{T}}(k)\,\mathbf{Q}_{xx}^{-1}(k-1)\,\mathbf{q}_{xr}(k-1) \\ &\quad + w(k)\,\mathbf{P}(k)\,\mathbf{x}_{\text{ch}}(k)\,r(k) \\ &= \hat{\mathbf{h}}(k-1) - \mathbf{K}(k)\,\mathbf{x}_{\text{ch}}^{\text{T}}(k)\,\hat{\mathbf{h}}(k-1) + w(k)\,\mathbf{P}(k)\,\mathbf{x}_{\text{ch}}(k)\,r(k)\end{aligned} \tag{5.49}$$

Using the fact that $w(k)\,\mathbf{P}(k)\,\mathbf{x}_{\text{ch}}(k)$ equals the gain vector $\mathbf{K}(k)$, as in eq. (5.47), we obtain the desired recursive equation for updating the tap–weight vector

$$\begin{aligned}\hat{\mathbf{h}}(k) &= \hat{\mathbf{h}}(k-1) + \mathbf{K}(k)\,\left[r(k) - \mathbf{x}_{\text{ch}}^{\text{T}}(k)\,\hat{\mathbf{h}}(k-1)\right] \\ &= \hat{\mathbf{h}}(k-1) + \xi(k)\,\mathbf{K}(k) \qquad (5.50) \\ &= \hat{\mathbf{h}}(k-1) + \mathbf{P}(k)\,\mathbf{x}_{\text{ch}}(k)\,w(k)\xi(k) \qquad (5.51)\end{aligned}$$

where $\xi(k)$ is the *a priori* estimation error defined in eq. (5.8).

Equations (5.8), (5.44), (5.50), and (5.45), collectively and in that order, constitute the *recursive weighted least squares* (RWLS) algorithm, summarised as follows:

Algorithm 5.5 (RWLS)

Initialisation: $\mathbf{P}(0) = \delta^{-1}\mathbf{I}, \quad \hat{\mathbf{h}}(0) = \mathbf{0}$

Basic recursion: $\forall\ k = 1, 2, \ldots$

1. *a priori* estimation error

$$\xi(k) = r(k) - \hat{\mathbf{h}}^{\mathrm{T}}(k-1)\,\mathbf{x}_{\mathrm{ch}}(k)$$

2. weighting sequence

$$w(k) = \phi(\xi(k))$$

3. gain vector

$$\mathbf{K}(k) = \frac{\lambda^{-1}\mathbf{P}(k-1)\mathbf{x}_{\mathrm{ch}}(k)}{w(k)^{-1} + \lambda^{-1}\mathbf{x}_{\mathrm{ch}}^{\mathrm{T}}(k)\mathbf{P}(k-1)\mathbf{x}_{\mathrm{ch}}(k)}$$

4. tap weight vector update

$$\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k-1) + \xi(k)\mathbf{K}(k)$$

5. inverse weighted autocorrelation matrix

$$\mathbf{P}(k) = \lambda^{-1}\mathbf{P}(k-1) - \lambda^{-1}\mathbf{K}(k)\mathbf{x}_{\mathrm{ch}}^{\mathrm{T}}(k)\mathbf{P}(k-1)$$

6. repeat from step 1

□

It is interesting to notice that the recursive equations of RWLS are almost identical to the traditional recursive least squares (RLS) algorithm, as presented in section 5.3. The only difference is the first term in the denominator of the gain vector $\mathbf{K}(k)$, i.e. $w(k)^{-1}$ instead of 1.

On the other hand, the same set of recursive equations can be derived from RLS using the generalised orthogonality criterion of eq. (5.30). Moreover, eq. (5.47) implies

$$\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k-1) + \mathbf{P}(k)\mathbf{x}_{\mathrm{ch}}(k)\,\psi(\xi(k)) \tag{5.52}$$

This is essentially equivalent to the recursive Gauss–Newton algorithm for general criteria [111,

pages 96–98]. This algorithm can be formulated for the channel estimation problem as follows

$$\mathbf{R}(k) = \mathbf{R}(k-1) + \gamma(k)\left[\mathbf{x}_{\text{ch}}(k)\rho''(\xi(k))\mathbf{x}_{\text{ch}}^{\text{T}}(k) - \mathbf{R}(k-1)\right] \tag{5.53}$$

$$\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k-1) + \gamma(k)\mathbf{R}^{-1}(k)\mathbf{x}_{\text{ch}}(k)\rho'(\xi(k)) \tag{5.54}$$

where $\gamma(k)$ is the *gain–sequence* of the Robins–Monro scheme. Observe that for a polynomial penalty function $\rho(x)$, its second derivative $\rho''(x)$ is proportional to $\phi(x)$ (see eq. (5.28)). If $\gamma(k) = 1/k$ then eq. (5.53) actually provides an estimate of the time–averaged weighted autocorrelation matrix $\mathbf{Q}_{xx}(k)/k$. Morover, eq. (5.54) is equivalent to the tap–weight adaptation formula in eq. (5.51). To recapitulate, we have shown that RWLS is a recursive approximation to the M–estimation problem, and furthermore, equivalent to the Gauss–Newton recursive algorithm.

5.5.1 Traditional recursive least squares (RLS)

The RWLS recursion is reduced to the traditional RLS algorithm if a quadratic penalty function is chosen, corresponding to

$$\rho_{LS}(x) = \frac{x^2}{2}, \quad \psi_{LS}(x) = x, \quad \phi_{LS}(x) = 1 \tag{5.55}$$

5.5.2 Recursive maximum likelihood (RML)

Assume that the noise samples are independent identically distributed (i.i.d.), drawn from a random variable with a stable distribution f_α. Then, the likelihood function of the received vector $\mathbf{r}$ under the parameters $\hat{\mathbf{h}}$ is given by

$$\mathcal{L}_{\hat{\mathbf{h}}}(\mathbf{r}; f_\alpha) \stackrel{\triangle}{=} \log \prod_{i=0}^{k} f_\alpha(e(i)) = \sum_{i=0}^{k} \log f_\alpha(e(i)) \tag{5.56}$$

Therefore, the maximum likelihood estimate (ML) $\hat{\mathbf{h}}$ is obtained by RWLS when

$$\rho_{ML}(x) = -\log f_\alpha(x), \quad \psi_{ML}(x) = -\frac{f'_\alpha(x)}{f_\alpha(x)}, \quad \phi_{ML}(x) = -\frac{f'_\alpha(x)}{x f_\alpha(x)} \tag{5.57}$$

The RML, however, is not of practical interest, since it requires that the pdf of the noise and its derivative are evaluated at each iteration. But, in general, there is no closed form expression

for the *S*α*S* pdf. For our simulations in section 5.7, numerical approximations of the *S*α*S* distribution have been used.

5.5.3 Recursive least p–norm (RLP)

As shown in [9] and discussed in section 3.6, the *minimum dispersion* criterion is a natural and mathematically meaningful choice as a measure of optimality in stable signal processing. Recall, that by minimising the error dispersion, we minimise the average magnitude of the estimation error. Furthermore, minimising the dispersion is also equivalent to minimising the probability of large estimation errors [70]. Finally, the norm of a *S*α*S* random variable is proportional to its p–th order moment for any $0 < p < \alpha$. Consequently, the appropriate cost function should be

$$\mathcal{J}_{LP} = \sum_{i=0}^{k} |e(i)|^p \tag{5.58}$$

and, therefore

$$\rho_{LP}(x) = \frac{|x|^p}{p}, \quad \psi_{LP}(x) = \langle x \rangle^{p-1}, \quad \phi_{LP}(x) = |x|^{p-2} \tag{5.59}$$

The choice of the weighting sequence $w(i)$ from eq. (5.59) for RWLS gives rise to the *recursive least p–norm* (RLP) algorithm for adaptive filtering in *S*α*S* noise environment. As shown in [9], all lower order moments of a *S*α*S* random variable are equivalent, i.e., the p–th and q–th order moments differ by a constant factor independent of the *S*α*S* random variable for any $0 < p < \alpha$ and $0 < q < \alpha$. This important result implies that, according to eq. (5.59), one does not have to explicitly estimate the actual noise characteristic exponent, but only set parameter p to be less than α.

Another approach to the least p–norm optimisation problem was taken by Byrd and Payne [112] in the form of the *iteratively re–weighted least squares* (IRLS) algorithm. Suppose that the $(k+1)$–vector $\mathbf{r}_k$ contains all the received samples from time 0 until time k. We now construct a $(k+1 \times N)$ matrix $\mathbf{X}_k$ such that its i–th row contains the transpose of the channel

input vector $\mathbf{x}_{\text{ch}}(i)$ at time i, i.e.,

$$\mathbf{X}_k = \begin{bmatrix} x(0) & 0 & \cdots & 0 \\ x(1) & x(0) & \cdots & 0 \\ & \vdots & & \vdots \\ x(k) & x(k-1) & \cdots & x(k-N+1) \end{bmatrix} \tag{5.60}$$

The traditional LS estimate can then be written as

$$\hat{\mathbf{h}}_{\text{LS}}(k) = \left(\mathbf{X}_k^{\text{T}}\mathbf{X}_k\right)^{-1}\ \mathbf{X}_k^{\text{T}}\mathbf{r}_k \tag{5.61}$$

Note that $\mathbf{X}_k^{\text{T}}\mathbf{X}_k$ is actually the sample autocorrelation matrix of the input $\mathbf{r}_{xx}(k)$ (eq. (5.18)), while $\mathbf{X}_k^{\text{T}}\mathbf{r}_k$ is the sample cross–correlation vector $\mathbf{r}_{xr}(k)$ (eq. (5.19)). The IRLS algorithm, on the other hand, solves the least p–norm problem by iteratively computing

$$\hat{\mathbf{h}}(j) = \left(\mathbf{X}_k^{\text{T}}\mathbf{W}(j-1)\mathbf{X}_k\right)^{-1}\ \mathbf{X}_k^{\text{T}}\mathbf{W}(j-1)\mathbf{r}_k \tag{5.62}$$

until the rate of change for the norm of the estimation error vector falls below a certain threshold ϵ, i.e,

$$\frac{\|\mathbf{e}(j)\|_p - \|\mathbf{e}(j-1)\|_p}{\|\mathbf{e}(j-1)\|_p} < \epsilon \tag{5.63}$$

where $\mathbf{e}(j)$ is the estimation error vector at the j–th iteration

$$\mathbf{e}(j) = \mathbf{r}_k - \mathbf{X}_k\hat{\mathbf{h}(j)} \tag{5.64}$$

Here, $\mathbf{W}(j)$ is a diagonal matrix with its diagonal elements defined as

$$\mathbf{W}_{ii}(k) = \phi(\mathbf{e}_i(j)) \tag{5.65}$$

where ϕ is the least p–norm weighting function (eq. (5.59)). A comprehensive discussion of least p–norm deconvolution and the IRLS algorithm is given in [72], while in [105] the authors use IRLS for non–linear autoregressive modelling.

The IRLS algorithm, although not optimal, converges to the least p–norm solution when the weighting function satisfies some fairly weak conditions. In particular, $\phi(x)$ must be non–

increasing in $|x|$, and bounded for all x [106, 112]. Furthermore, IRLS produces estimates which are closer to the optimum least p–norm solution compared to the recursive least p–norm (RLP) algorithm. This is because the elements of the residual vector $\mathbf{e}(j)$ converge, at every iteration, to the sequence $v(i)$ (eq. (5.34)) which corresponds to the estimation error for the optimal least p–norm estimate $\hat{\mathbf{h}}_{\text{opt}}(k)$ at time k.

The role of IRLS, however, in a communications signal processing context is questionable, due to a number of reasons. First of all, its iterative process has to be repeated for every new data sample that arrives. Secondly, the adequate number of iterations for the algorithm to converge is not known beforehand. That makes the algorithm unsuitable for real–time applications. But, more importantly, IRLS requires infinitely growing storage memory, because the matrices $\mathbf{X}_k$, $\mathbf{W}(j)$ and vector $\mathbf{r}_k$ grow as new data arrive. Finally, the computational complexity grows as well (in a square law manner) due to the growth of the matrices. On the other hand, the RLP algorithm, although less optimal than IRLS, is a purely recursive algorithm and is characterised by constant requirements in computational complexity and storage space, which do not grow with time.

Equation (5.59) suggests that, in general, $w(i)$ is not bounded. That is, for infinitesimally small error the corresponding weight is infinitely large. The theoretical justifications in [106, 112] however, require a bounded weighting sequence. Moreover, considering eq. (5.39) and eq. (5.40), infinitely large terms may result to numerical instability in the computation of $\mathbf{Q}_{xx}$ and $\mathbf{q}_{xr}$. Huber [106], suggested that it is desirable to bound the sequence $w(i)$ for very small samples of the estimation error, as follows

$$w(i) = \begin{cases} |\xi(i)|^{p-2} & , \quad |\xi(i)| > \omega \\ \omega^{p-2} & , \quad |\xi(i)| \leq \omega \end{cases} \tag{5.66}$$

where ω is a small positive constant. Essentially, for samples with small estimation error (where probably the noise term is very small) we use the traditional LS penalty function ($\rho(x) = x^2/2$).

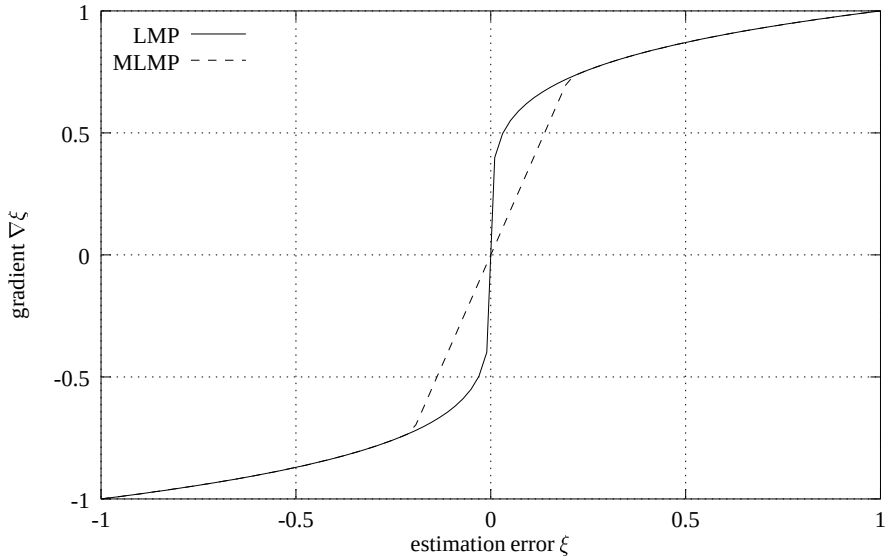


Figure 5.6: *Estimation error gradient for LMP and MLMP ($\alpha = 1.2$).*

5.6 Modified LMP

Since $\langle \xi(k) \rangle^{p-1} = |\xi(k)|^{p-2}\, \xi(k)$ we can rewrite the tap–weight adaptation formula of LMP (eq. (5.11)) as

$$\hat{\mathbf{h}}(k+1) = \hat{\mathbf{h}}(k) + \mu\, |\xi(k)|^{p-2}\, \xi(k)\, \mathbf{x}_{\text{ch}}(k) \tag{5.67}$$

Comparing now eq. (5.67) to eq. (5.51) it follows that the LMP algorithm is actually a reduced formulation of RLP, with a constant $\mathbf{P}(k) = \mu\mathbf{I}$, and an unbounded weighting function $|\xi(k)|^{p-2}$.

This results to a steep estimation error gradient close to zero (e.g. fig. 5.6 – solid line), making LMP more sensitive to gradient noise in comparison to the conventional LMS. Replacing $|\xi(k)|^{p-2}$ in eq. (5.67) with the bounded weighting sequence $w(i)$ from eq. (5.66) we obtain a stochastic gradient algorithm with a less steep gradient close to zero (fig. 5.6 – dashed line) and therefore less misadjustment.

For a real–time channel estimation system we could also employ a time varying step–size parameter $\mu(k) = (1 + c\nu^k)\, \mu_0$ in order to speed–up the transient behaviour of the mean–square error. The step–size parameter starts equal to $(1 + c)\mu_0$ and, following an exponential decline, reaches asymptotically the value μ_0 (see fig. 5.7). Here c is a scale constant for the initial value of $\mu(k)$ and ν ($0 < \nu < 1$) is the exponential rate, controlling the speed of the transient of

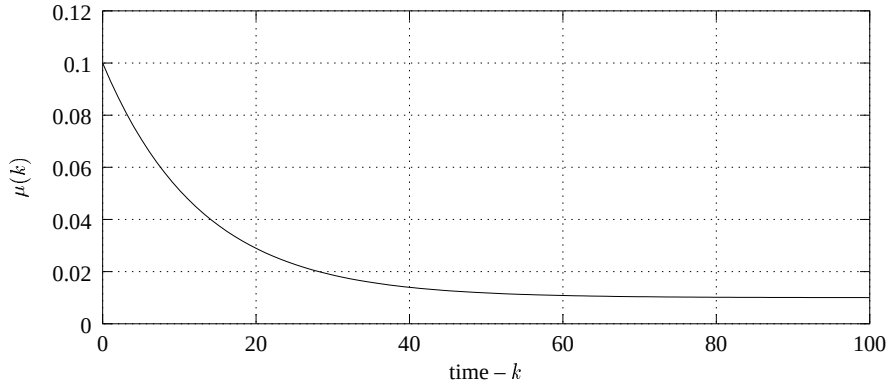


Figure 5.7: *Time varying $\mu(k)$ with $\nu = 0.925$, $c = 9$ and $\mu_0 = 0.01$.*

the step–size parameter. We can, therefore, summarise a modified least mean p–norm (MLMP) algorithm

Algorithm 5.6 (MLMP) Fix p so that $1 \leq p < \alpha$.

Initialisation: $\hat{\mathbf{h}}(0) = \mathbf{0}$

Basic recursion: $\forall\ k = 1, 2, \ldots$

1. *a priori* estimation error

$$\xi(k) = r(k) - \hat{\mathbf{h}}^{\mathrm{T}}(k-1)\mathbf{x}_{\mathrm{ch}}(k)$$

2. weighting sequence

$$w(k) = \begin{cases} |\xi(k)|^{p-2} & , \quad |\xi(k)| > \omega \\ \omega^{p-2} & , \quad |\xi(k)| \leq \omega \end{cases}$$

3. tap weight adaptation

$$\hat{\mathbf{h}}(k+1) = \hat{\mathbf{h}}(k) + (1 + c\nu^k)\ \mu_0\ w(k)\ \xi(k)\ \mathbf{x}_{\mathrm{ch}}(k) \tag{5.68}$$

4. repeat from step 1

□

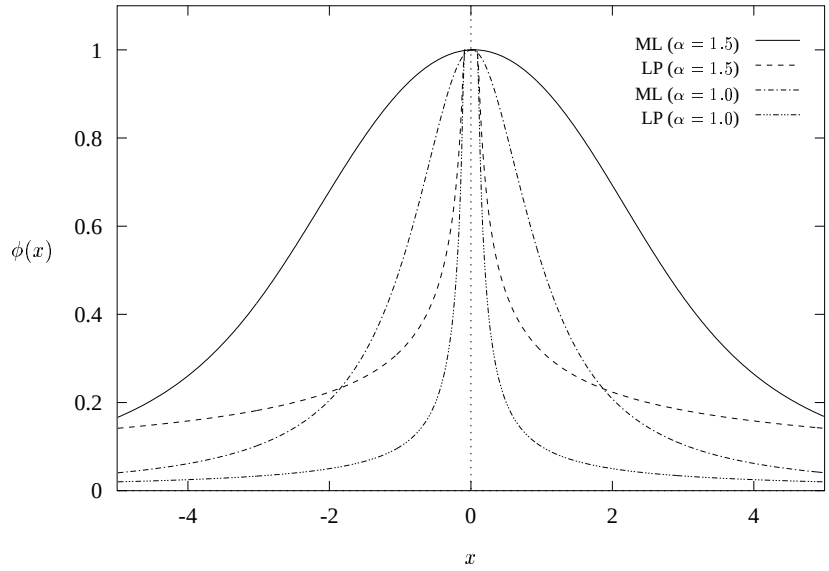


Figure 5.8: *Weighting functions $\phi(x)$ for Maximum Likelihood and Least p–norm criteria for two values of the characteristic exponent α.*

5.7 Experiments

In fig. 5.8 we plot the weighting functions ϕ for the two non–linear optimisation criteria (ML and LP), for the cases $\alpha = 1.5$ and $\alpha = 1$. Note that, the larger the magnitude of the residuals, the smaller the weight these criteria produce. Indeed, large residuals are likely to be the result of noise impulses, and hence, are weighted less heavily in the cost function.

Unfortunately there is no convergence and stability analysis for LMP or RWLS. Nevertheless, the experimental data suggest that the algorithms converge efficiently and produce satisfactory estimates of the channel impulse response in impulsive noise environments. Our experiments for channel identification using the algorithms LMP, MLMP, OSRLS, RML and RLP have been carried out with channel impulse response

$$\mathbf{h} = [\ .04 \quad -.05 \quad .07 \quad -.21 \quad -.5 \quad .72 \quad .36 \quad 0 \quad .21 \quad .03 \quad .07\ ]^{\mathrm{T}} \tag{5.69}$$

and noise parameters $\alpha = 1, \quad \gamma = .08, \quad \delta = 0$. The step–size parameter for both LMP and MLMP was $\mu = .004$. For MLMP $\mu_0 = .004$, $\nu = 0.875$ and $c = 10$. For OSRLS, $\eta = 3$ and $\Theta = 9$. The forgetting factor λ for RWLS was set to 0.98. Finally, the constant ω for the least p–norm bounded weighting sequence (eq. (5.66)) was chosen 0.1 (for MLMP and RLP).

Figure 5.9 depicts the ensemble (over 400 Monte–Carlo runs) mean squared error for algorithms

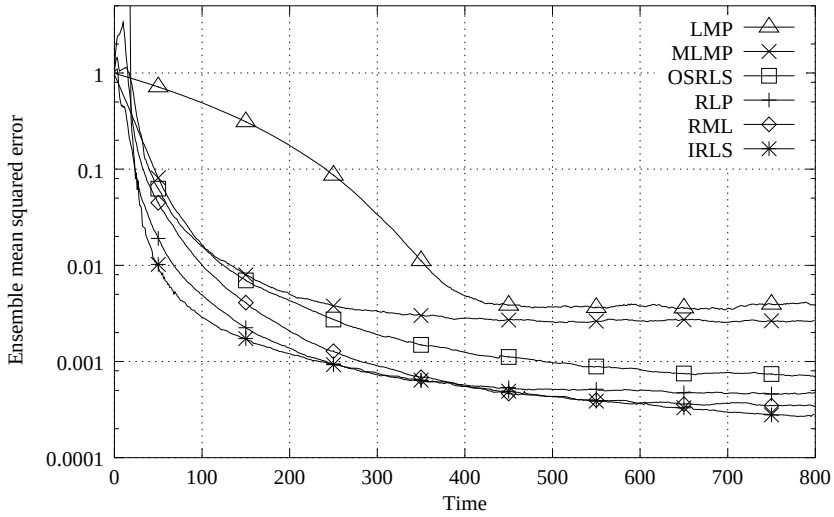


Figure 5.9: *The convergence of LMP, MLMP, OSRLS, RML, RLP and IRLS, for a channel with 11 taps.*

LMP, MLMP, OSRLS, RML and RLP. The ensemble convergence of IRLS for the same experimental setup is also depicted in order to obtain a relative measure for the performance of the recursive algorithms, since IRLS offers the best known performance for the least p–norm optimisation problem.

Clearly, all LS type algorithms outperform the stochastic gradient algorithms. The convergence of MLMP is better than LMP in terms of both transient behaviour and steady state misadjustment. In fact, the transient convergence of MLMP is comparable with that of LS algorithms, since the input is white. The enhanced transient behaviour of MLMP is due to the time–varying step–size parameter. On the other hand, the steady state misadjustment of MLMP is better than LMP, although both algorithms were simulated with the same steady–state step–size parameter. This improvement should be accredited to the bounded weighting sequence which provides less sensitivity to the gradient noise.

Among the recursive LS type algorithms (i.e., OSRLS, RLP and RML), RML achieves, as expected, the best asymptotic performance. However, RML and RLP do not retain the defining characteristic of the traditional LS scheme, i.e., that the mean squared estimation error continuously deminishes as $k \to \infty$. This behaviour can be found in IRLS. On the contrary, the mean squared error of RML and RLP seems to reach an asymptotic infimum, a behaviour similar to the stochastic gradient algorithms. However, this infimum is significantly lower than LMP or MLMP. The transient behaviour of RLP is superior to RML, and actually comparable to that of IRLS. Finally, the performance of OSRLS is poorer than RML and RLP because, for the

specific values of η and Θ, this algorithm discards about a third of the received samples.

In summary, IRLS offers the best known performance for least p–norm optimisation, but at an unaffordably high computational cost. Alternatively, there is a variety of recursive algorithms with reasonable complexity but compromised performance. Among these, the most suitable for channel estimation in a receiver are MLMP and RLP. They are both direct generalisations of the conventional LMP and RLS, respectively, with negligible extra computational requirements, providing robust performance in impulsive non–Gaussian environments. In particular, RLP achieves a performance which is remarkably close to that of IRLS, but at a considerably lower computational complexity.

5.8 The clustering technique

An alternative strategy for estimating the centers of the equaliser is the clustering approach. Recall from chapter 4, that the received vector $\mathbf{r}(k)$ forms α–stable clusters around the noise–free centers $\mathbf{c}_i$ of the channel. This suggests that a clustering (or arithmetic averaging) procedure should be able to filter out the noise and produce an estimate of the equaliser centers.

In early studies of the clustering technique [76], the algorithm was designed to estimate the 2^{N+M-1} equaliser centers $\mathbf{c}_i$ ($0 \leq i < N_{\text{c}}$), directly from the received vector $\mathbf{r}(k)$. However, as shown in [33], it is only sufficient to estimate the 2^N scalar channel states $\bar{c}_i$ ($0 \leq i < N_{\text{sc}}$). Consequently, the vector centers $\mathbf{c}_i$ of the equaliser can be derived from these states through eq. (4.16) (page 46).

In order to estimate the channel noise-free states, the receiver first needs to identify which cluster each received sample $r(k)$ belongs to. During the training period, the transmitted symbols (and consequently the channel input vector $\mathbf{x}_{\text{ch}}$) are known to the receiver (supervised learning). The relation for the channel states was given in eq. (4.8). According to this relation, the channel states $\bar{c}_i$ are uniquely determined by the channel input vector $\mathbf{x}_{\text{ch}}$. Therefore, it is necessary to define a mapping from all 2^N possible states of the channel input vector (referred to as $\mathbf{x}_{\text{ch}_i}$) to the appropriate channel state. One such mapping is

$$\mathbf{x}_{\text{ch}_i} \rightarrow \bar{c}_s \quad : \quad s = \left\lfloor \frac{\mathbf{b}^{\text{T}}\mathbf{x}_{\text{ch}_i} + 2^N - 1}{2} \right\rfloor \tag{5.70}$$

where

$$\mathbf{b} = [2^0 \quad 2^1 \quad 2^2 \quad \ldots \quad 2^{N-1}]^{\mathrm{T}} \tag{5.71}$$

is a N–vector containing, in ascending order, all powers of the form 2^i ($0 \leq i < N$). The transformation of eq. (5.70) produces an integer index s, such that $0 \leq s < 2^N$.

So, at each time instant k, it can be inferred from the channel input vector $\mathbf{x}_{\mathrm{ch}}(k)$ which channel state $\bar{c}_i$ occurs. The clustering scheme, then, averages all the received samples associated to a specific channel state $\bar{c}_i$ in order to extract an estimate $\hat{\bar{c}}_i$ for this state. The computational procedure of this clustering algorithm can be summarised as follows:

Algorithm 5.7 (Supervised clustering)
Initialisation: $\hat{\bar{c}}_i(k) = 0 \quad , \quad m_i = 0 \quad : \quad \forall\, i = 0,\, 1,\, \ldots,\, N_{\mathrm{sc}}$
Basic recursion: $\forall\, k = 1, 2, \ldots$

1. identify the channel state

$$s = \left\lfloor \frac{\mathbf{b}^{\mathrm{T}} \mathbf{x}_{\mathrm{ch}}(k) + 2^N - 1}{2} \right\rfloor$$

2. update channel state estimate

$$\hat{\bar{c}}_s(k) \quad = \quad \frac{m_s\, \hat{\bar{c}}_s(k-1) + r(k)}{m_s + 1}$$

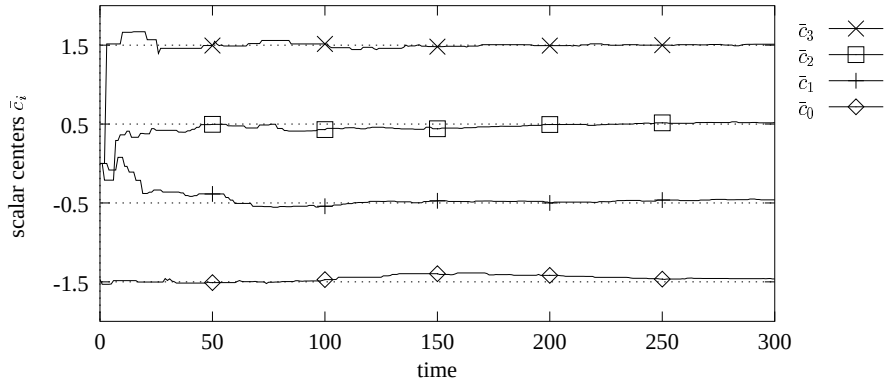
3. increment the s–th counter

$$m_s = m_s + 1$$

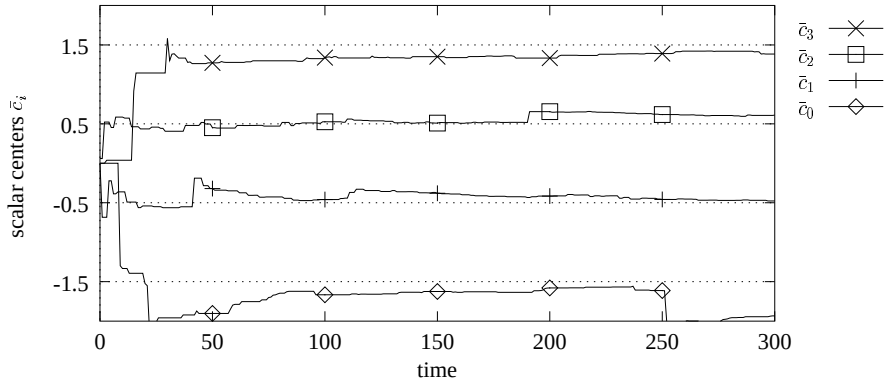
4. repeat from step 1

□

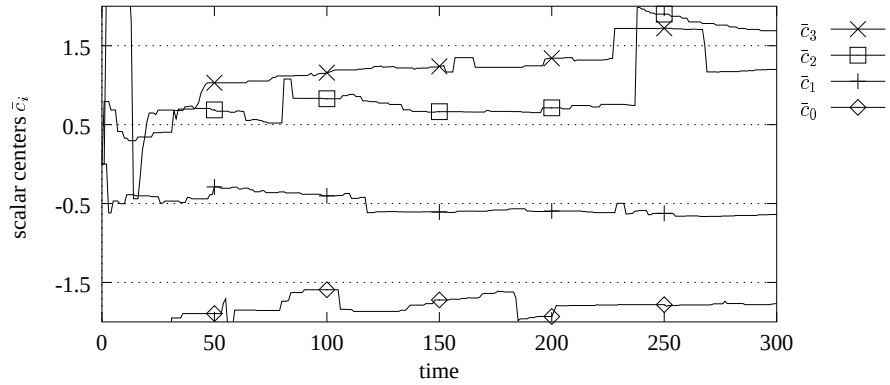
In a Gaussian noise environment, this clustering technique converges efficiently to the scalar channel states. For example, fig. 5.10(a) depicts the convergence of the channel states estimates for channel $H(z) = 1 + 0.5z^{-1}$. This channel has 4 noise–free states, namely -1.5, -0.5, 0.5, and 1.5. However, when the noise statistics deviate from the pure normal distribution, the performance of the algorithm is degraded significantly, due to the frequent occurrence of large



(a) Gaussian noise ($\alpha = 2$, $\gamma = 0.25$).



(b) α–stable noise ($\alpha = 1.5$, $\gamma = 0.25$).



(c) α–stable noise ($\alpha = 1$, $\gamma = 0.25$).

Figure 5.10: *The convergence of supervised clustering for three different noise environments.*

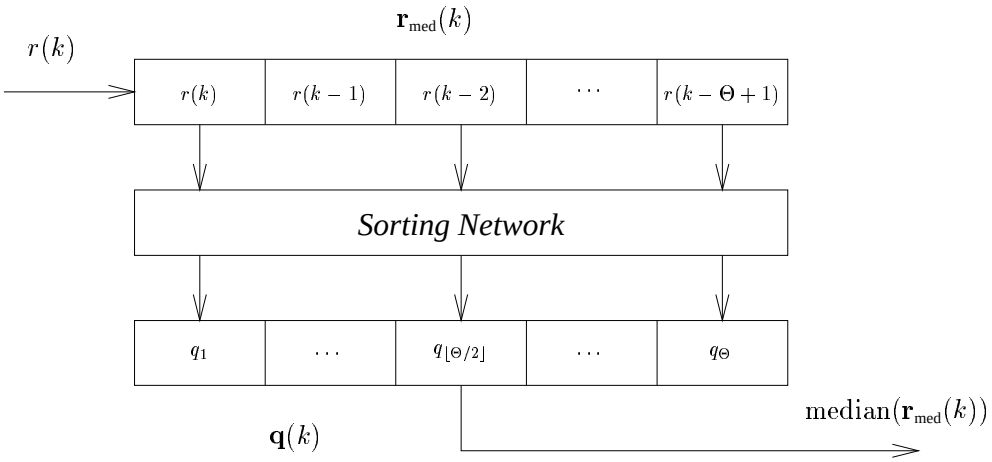


Figure 5.11: *The median filter.*

interference levels. For example figures 5.10(b) and 5.10(c) depict the convergence of the state estimates for $\alpha = 1.5$ and $\alpha = 1$, respectively.

The clustering technique can be made robust against the outliers, by introducing a median filter at the input to algorithm 5.7. The median filter (depicted in fig. 5.11) is well known for rejecting the impulsive noise often experienced in image processing [113]. In general, it can effectively discard the outliers in a data set, but its use in communications signal processing is questionable, due to the fact that it does not preserve the time sequence of the data. However, for the purpose of estimating the channel states, it can be found useful.

Specifically, a number Θ of the most recent received data $r(k)$ are stored in vector $\mathbf{r}_{\text{med}}$. This vector is then sorted, and the $\lfloor\Theta/2\rfloor$–th in order element of the resulting vector $\mathbf{q}$ is chosen as the median of the data contained in $\mathbf{r}_{\text{med}}(k)$. In this way, the samples with large noise magnitude are directed towards the edges of the sorted vector $\mathbf{q}$, while the central elements are close to the location of the distribution of the data. A highly desirable characteristic of the median filter is that the only adjustable parameter is its order Θ.

Special care must be taken, however, in order to integrate a median filter in the clustering scheme. This is due to the fact that the median filter does not preserve the time sequence of the data. Thus, separate vectors $\mathbf{r}^i_{\text{med}}$ ($0 \leq i < N_{\text{sc}}$) should be employed to store the Θ most recent received samples associated with a specific channel state $\bar{c}_i$. Then, the output from each median filter passes through an arithmetic averaging procedure, to produce an estimate of the corresponding channel state. The block diagram of this scheme is depicted in fig. 5.12. This scheme will be referred to as *median clustering*, and the algorithm can be summarised as

follows

Algorithm 5.8 (Supervised median clustering)

Initialisation: $\hat{\tilde{c}}_i(k) = 0 \quad , \quad m_i = 0 \quad : \quad \forall\, i = 0,\, 1,\, \ldots\, ,\, N_{\text{sc}}$

Basic recursion: $\forall\, k = 1, 2, \ldots$

1. identify the channel state

$$s = \left\lfloor \frac{\mathbf{b}^{\text{T}}\mathbf{x}_{\text{ch}}(k) + 2^N - 1}{2} \right\rfloor$$

2. insert the received sample $r(k)$ into the appropriate median vector $\mathbf{r}^s_{\text{med}}$.

3. update channel state estimate

$$\hat{\tilde{c}}_s(k) \quad = \quad \frac{m_s\, \hat{\tilde{c}}_s(k-1) + \text{median}(\mathbf{r}^s_{\text{med}}(k))}{m_s + 1}$$

4. increment the s–th counter

$$m_s = m_s + 1$$

5. repeat from step 1

□

As shown in figures 5.13(a), 5.13(b), and 5.13(c), the use of the median filter improves significantly the performance of the clustering technique in impulsive noise environments, compared to figure fig. 5.10. For this experiment, the order of the median filter was $\Theta = 3$.

A disadvantage of this technique, though, is that it does not produce meaningful estimates before all median vectors $\mathbf{r}^i_{\text{med}}$ have been filled with received samples. If the training sequence is chosen carefully so that all channel states occur with the same frequency, this period lasts $\Theta 2^N$ samples. The main drawback of the median clustering scheme, however, is that there is no analytical way of determining the order of the filter Θ, but it can only be set heuristically.

The clustering algorithms 5.7 and 5.8 both employ an arithmetic averaging in order to filter out the noise of the received samples. While this method is adequate for stationary channels, it can not track the states of a time–varying channel. An alternative version of the clustering

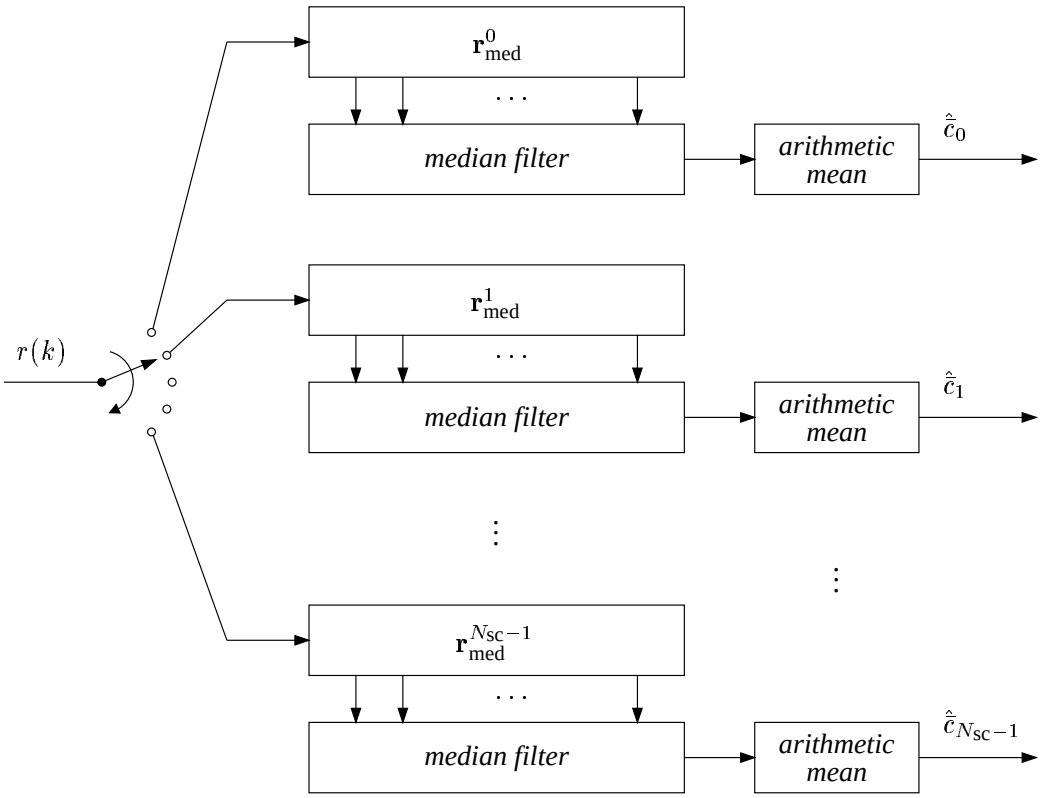


Figure 5.12: *Block diagram of multiple median filters for robust clustering.*

algorithm 5.7 for non–stationary scenarios is proposed in [76], and can be summarised in the following lines

Algorithm 5.9 (Supervised clustering for non–stationary environments)

Initialisation: $\hat{\bar{c}}_i(k) = 0 \quad : \quad \forall\, i = 0,\ 1,\ \ldots,\ N_{\text{sc}}$

Basic recursion: $\forall\ k = 1, 2, \ldots$

1. identify the channel state

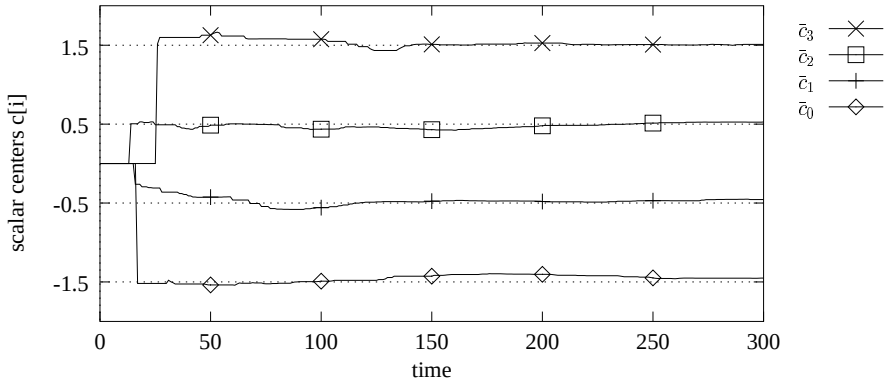
$$s = \left\lfloor \frac{\mathbf{b}^{\text{T}}\mathbf{x}_{\text{ch}}(k) + 2^N - 1}{2} \right\rfloor$$

2. update channel state estimate

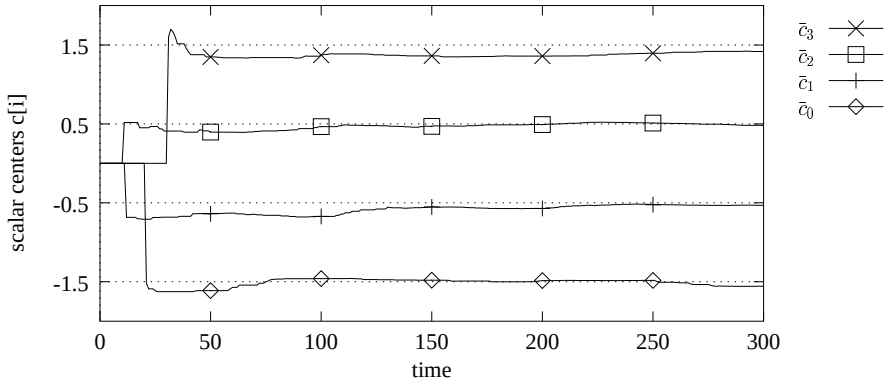
$$\hat{\bar{c}}_s(k) \quad = \quad \hat{\bar{c}}_s(k-1) + \mu\left(r(k) - \hat{\bar{c}}_s(k)\right)$$

3. repeat from step 1

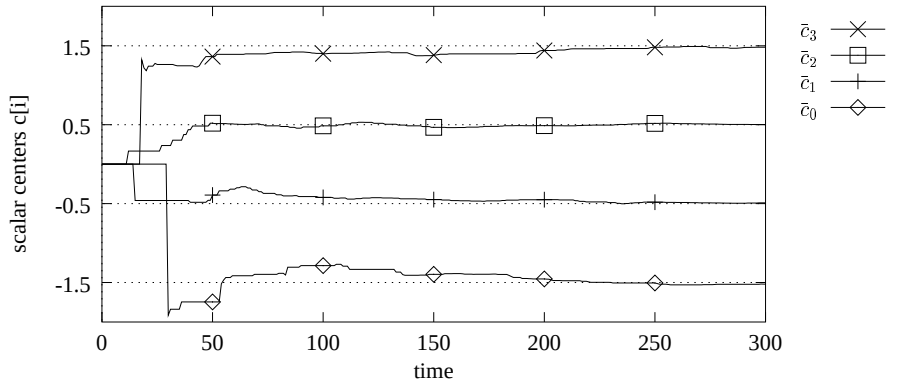
□



(a) Gaussian noise ($\alpha = 2$, $\gamma = 0.25$).



(b) α–stable noise ($\alpha = 1.5$, $\gamma = 0.25$).



(c) α–stable noise ($\alpha = 1$, $\gamma = 0.25$).

Figure 5.13: *The performance of supervised median clustering for a variety of values for the noise characteristic exponent* α.

where, μ is a small positive constant, the *learning rate*. Accordingly, it is straightforward to modify the median clustering algorithm in order to adapt to non–stationary environments

Algorithm 5.10 (Supervised median clustering for non–stationary environments)
Initialisation: $\hat{\bar{c}}_i(k) = 0 \quad : \quad \forall\, i = 0,\ 1,\ \ldots,\ N_{\text{sc}}$
Basic recursion: $\forall\, k = 1, 2, \ldots$

1. identify the channel state

$$s = \left\lfloor \frac{\mathbf{b}^{\text{T}}\mathbf{x}_{\text{ch}}(k) + 2^N - 1}{2} \right\rfloor$$

2. insert the received sample $r(k)$ into the appropriate median vector $\mathbf{r}^s_{\text{med}}$.

3. update channel state estimate

$$\hat{\bar{c}}_s(k) = \hat{\bar{c}}_s(k) + \mu \left(\text{median}\left(\mathbf{r}^s_{\text{med}}(k)\right) - \hat{\bar{c}}_s(k) \right)$$

4. repeat from step 1

□

A closer examination of the adaptive clustering algorithms 5.9 and 5.10 reveals that they do not employ counters for the calculation of the channel states estimates $\hat{\bar{c}}_s$, in contrast with algorithms 5.7 and 5.8. This characteristic offers an extra advantage to the adaptive versions of the algorithms. Indeed, since these counters are not bounded, the implementation of algorithms 5.7 and 5.8 is not of practical interest.

The tracking performance of both (traditional and median) adaptive clustering algorithms in an impulsive noise environment was simulated. The channel was time–varying with 2 coefficients, and a transfer function

$$H(z) = 1 + (0.2 + 0.0015\, k)\, z^{-1}$$

where k is the discrete time index. The characteristic exponent was chosen $\alpha = 1$. The order of the median filter was 3. The center trajectories obtained using the two adaptive clustering algorithms with a learning rate $\mu = 0.2$, are plotted in fig. 5.14 (for the traditional adaptive clustering algorithm) and fig. 5.15 (for the median adaptive clustering algorithm). The true

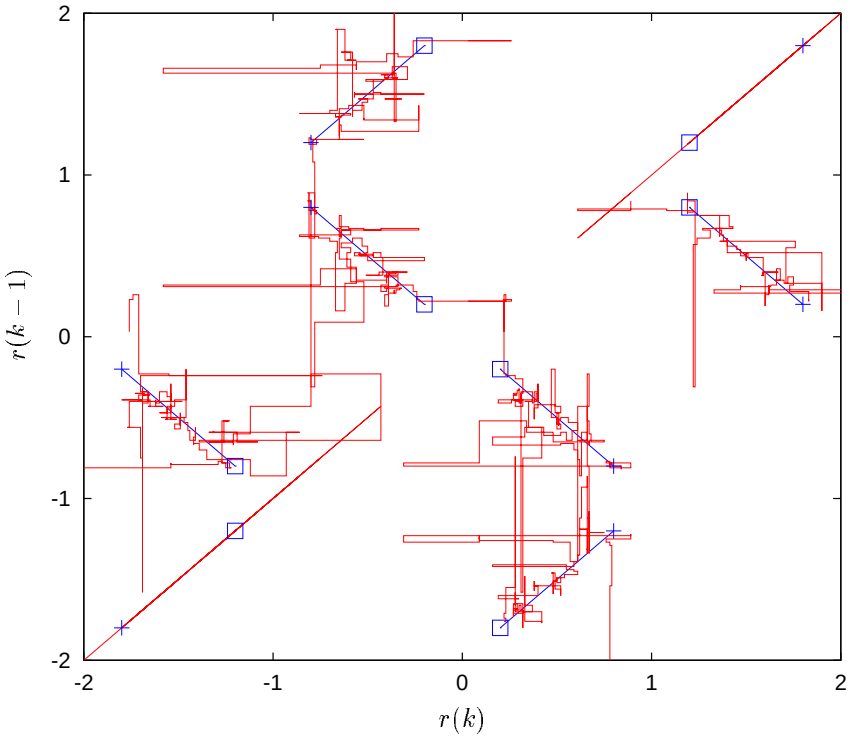


Figure 5.14: *Tracking performance of the traditional adaptive clustering in impulsive noise environment.*

centers trajectories are depicted with ◇——+, while the estimated trajectories with ——. It is clearly seen that, in this highly impulsive noise environment, the tracking ability of the traditional clustering technique is extremely poor. This is because the algorithm diverges when large noises samples (outliers) occur frequently (fig. 5.14). On the other hand, the adaptive median clustering algorithm is robust in this highly impulsive noise environment, and its tracking performance is satisfactory (fig. 5.15).

5.8.1 Robust location estimation

Although median clustering is a robust technique for estimating the channel states in an impulsive noise environment, it is not efficient in terms of computational complexity and memory requirements. In this section we present a novel algorithm for estimating the channel states, designed under the light of the findings in section 5.5. First, we note that all the received samples associated with a certain channel state $\hat{c}_i$ are distributed with the same density as the noise, located at the corresponding channel state. Therefore, the task is to design a location estimator for the subset of received data associated with a specific channel state.

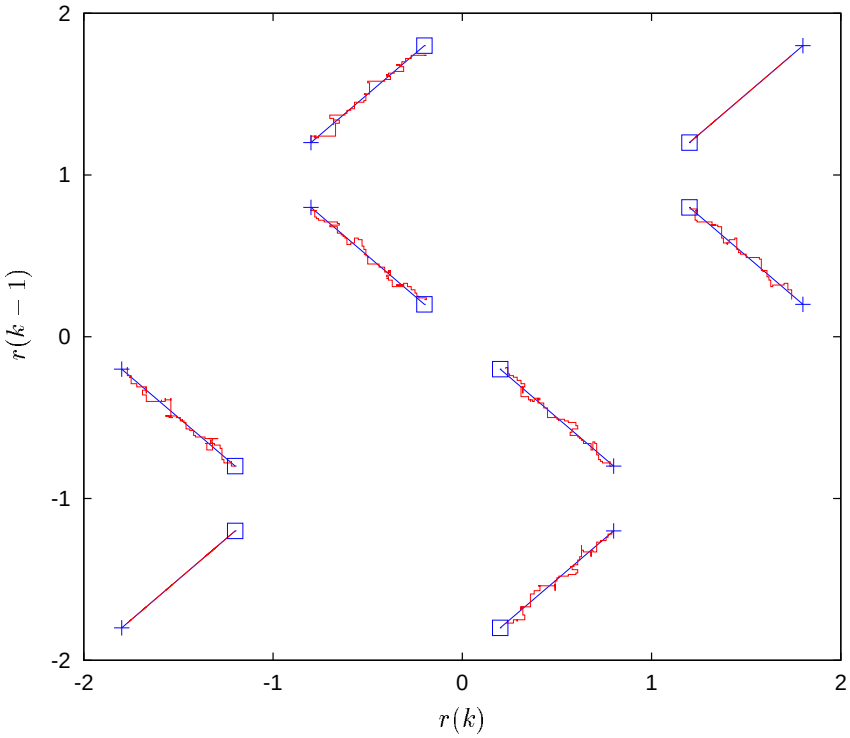


Figure 5.15: *Tracking performance of the adaptive median clustering in impulsive noise environment.*

Consider a non–stationary location $y(k)$ which is corrupted with additive noise $n(k)$, so that the observed signal is $r(k) = y(k) + n(k)$. Suppose that an estimate of the location is say $\hat{y}(k)$. The estimation error is then

$$e(k) = r(k) - \hat{y}(k) \tag{5.72}$$

The traditional least squares estimator minimises the weighted squared sum of estimation error samples

$$\mathcal{J}_{LS} = \sum_{i=0}^{k} \lambda^{k-i} (e(i))^2 \tag{5.73}$$

The parameter λ ($0 < \lambda \leq 1$) is the exponential weighting factor and corresponds to an equivalent window $W = 1/(1-\lambda)$ [35]. An M–estimator, on the other hand, would minimise

the weighted sum of a less rapidly increasing function ρ of the estimation error

$$\mathcal{J}_M = \sum_{i=0}^{k} \lambda^{k-i} \rho(e(i)) \tag{5.74}$$

Suppose that ρ has a derivative $\psi = \rho'$; then, the minimisation of eq. (5.74) implies

$$\begin{aligned}
\sum_{i=0}^{k} \lambda^{k-i} \psi(e(i)) = 0 \quad &\Leftrightarrow \\
\sum_{i=0}^{k} \lambda^{k-i} \phi(e(i)) e(i) = 0 \quad &\Leftrightarrow \\
\sum_{i=0}^{k} \lambda^{k-i} v(i) r(i) = \sum_{i=0}^{k} \lambda^{k-i} v(i) \hat{y}_{\mathrm{M}}(k) \quad &\Leftrightarrow \\
\hat{y}_{\mathrm{M}}(k) = \frac{\sum_{i=0}^{k} \lambda^{k-i} v(i) r(i)}{\sum_{i=0}^{k} \lambda^{k-i} v(i)} &
\end{aligned} \tag{5.75}$$

where $\phi(\cdot)$ and $v(i)$ are defined as in eq. (5.28). The weighting sequence $v(i)$ assumes knowledge of the optimum solution $\hat{y}_{\text{opt}}(k)$ at time k to generate the error sequence $e(i)$. Consequently, we will proceed as in section 5.5 by utilising the sequence $w(i) = \phi(\xi(i))$ in place of $v(i) = \phi(e(i))$, where $\xi(i) = r(i) - \hat{y}(i-1)$ is the *a priori* estimation error. Therefore, we obtain

$$\hat{y}_{\mathrm{M}}(k) = \frac{\sum_{i=0}^{k} \lambda^{k-i} w(i) r(i)}{\sum_{i=0}^{k} \lambda^{k-i} w(i)} \tag{5.76}$$

If we set

$$S_{\mathrm{M}}(k) = \sum_{i=0}^{k} \lambda^{k-i} w(i) \tag{5.77}$$

the recursive formula for $S_{\mathrm{M}}(k)$ is

$$\begin{aligned}
S_{\mathrm{M}}(k) &= w(k) + \lambda \sum_{i=0}^{k-1} \lambda^{k-1-i} w(i) \\
&= w(k) + \lambda S_{\mathrm{M}}(k-1)
\end{aligned} \tag{5.78}$$

The numerator in the right hand side of eq. (5.76) can be written

$$\begin{aligned}\sum_{i=0}^{k} \lambda^{k-i} w(i) r(i) &= w(i) r(k) + \lambda \sum_{i=0}^{k-1} \lambda^{k-1-i} w(i) r(i) \\ &= w(i) r(k) + \lambda S_{\mathrm{M}}(k-1) \hat{y}_{\mathrm{M}}(k-1) \end{aligned} \tag{5.79}$$

Hence, the M location estimator can be summarised as follows

Algorithm 5.11 (M location estimator)

Initialisation: $S_{\mathrm{M}}(0) = 0, \quad \hat{y}_{\mathrm{M}}(0) = 0$

Basic recursion: $\forall\ k = 1, 2, \ldots$

1. weighting sequence

$$w(k) = \phi(r(k) - \hat{y}_{\mathrm{M}}(k-1))$$

2. location estimate

$$\hat{y}_{\mathrm{M}}(k) = \frac{\lambda S_{\mathrm{M}}(k-1) \hat{y}_{\mathrm{M}}(k-1) + w(k) r(k)}{\lambda S_{\mathrm{M}}(k-1) + w(k)}$$

3. calculate $S_{\mathrm{M}}(k)$

$$S_{\mathrm{M}}(k) = \lambda S_{\mathrm{M}}(k-1) + w(k)$$

4. repeat from step 1

□

Note that if we set $\phi(x) = 1$, we obtain the traditional least squares estimator. Furthermore, setting $\lambda = 1$, the LS estimator actually becomes identical to the arithmetic mean of the samples. That is, the traditional clustering algorithm produces the LS estimate for the channel states with an infinite length window. On the other hand, choosing $\phi(x) = |x|^{p-2}$, we obtain a least p–norm location estimator, which is appropriate for any stable environments with $\alpha > p$. It is interesting to notice that for $p = 1$, the LP location estimator produces the median of the data set $\{r(k)\}$. However, it is not identical to the median clustering algorithm described in section 5.8.

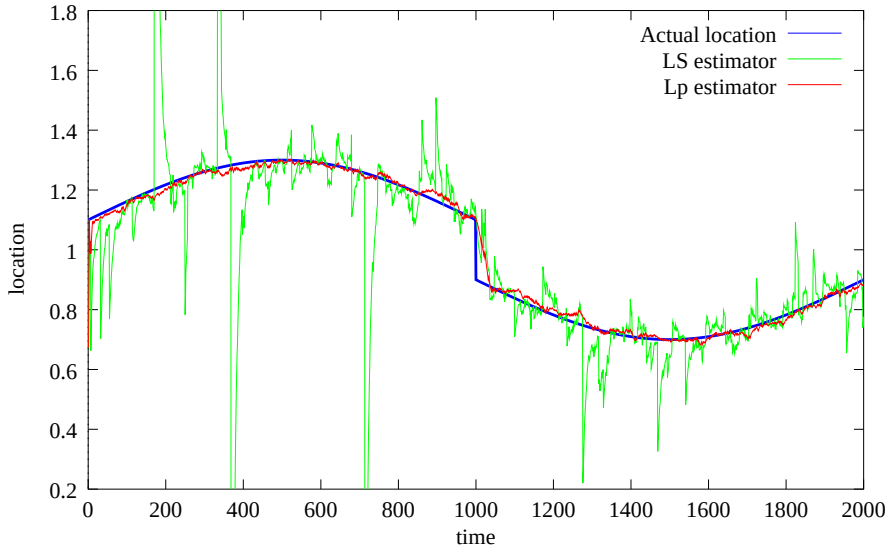


Figure 5.16: *Location estimation experiment with least squares and least p–norm algorithms.*

As an example, we consider the following non–stationary location

$$y(k) = 1 + 0.2\sin(2\pi fk) + 0.1\text{sgn}(\sin(2\pi fk)) \tag{5.80}$$

where $f = 5 \cdot 10^{-4}$. The trace of $y(k)$ is depicted in fig. 5.16 in blue, for 2000 samples. The location $y(k)$ is corrupted by $S\alpha S$ noise with characteristic exponent $\alpha = 1$, dispersion $\gamma = 0.04$, and location $\delta = 0$. Two location estimators were simulated, namely the least squares and the least p–norm. The p parameter for the latter was set equal to 1. The exponential factor λ was 0.9, which corresponds to an equivalent window $W = 10$. This choice was made in order to enable the estimators to track the non–stationary characteristics of $\{y(k)\}$. As fig. 5.16 shows, the LS estimate (in green) is extremely poor, because the noise is non–Gaussian. On the contrary, the LP estimate (in red) is robust in this highly impulsive noise environment and demonstrates a satisfactory tracking ability.

5.9 Estimation of the stable parameters

The stable distribution is fully determined by four parameters: the characteristic exponent α, the skew index β, the dispersion γ and the location parameter δ. An important problem is that of estimating these four parameters from observations of a stable random variable. For example, an adaptive channel equaliser operating in a stable noise environment, would be required to estimate the parameters of the noise from the actual received data. In most communication

systems the noise is symmetric around zero, so the assumption $\beta = 0$ and $\delta = 0$ is reasonable.

For the estimation of α and γ, most of the conventional methods in mathematical statistics, including the maximum likelihood estimation method, can not be used since these techniques depend on an explicit form for the noise distribution. However, a number of suboptimal numerical techniques have been developed that have been found useful in practice. In this section we introduce some of these techniques. An overview of these methods can be found in [9], while a more detailed presentation of the algorithms accompanied with experimental evaluation is given in [58]. The method of approximate maximum likelihood developed by DuMouchel [114] and evaluated by Brorsen and Yang [115] is omitted because it involves a highly nonlinear optimisation problem and no initialisation and convergence analysis is available.

Suppose the RV X is drawn from a stable distribution with parameters α, β, γ and δ. As shown in eq. (3.11), the RV $Y = (X - \delta)/\gamma^{\frac{1}{\alpha}}$ is α–stable with the same characteristic exponent and skew index, but with dispersion 1 and location 0. For convenience, the parameters estimation algorithms use the term $c = \gamma^{\frac{1}{\alpha}}$ instead of the dispersion parameter γ. Thus, a multiplicative factor in a random variable scales the parameter c in a linear manner.

5.9.1 Quantile based techniques

Suppose a random variable X has a continuous real–valued probability density function $f(x)$. The p–th quantile of $f(x)$ is the number x_p below which a fraction p of the values of X is expected to lie. That is, if the cumulative density function of X is $F(x)$, then x_p satisfies

$$F(x_p) = p \tag{5.81}$$

Thus, x_p is the inverse of the function $p = F(x)$ [38]. Given a set of N observations, a consistent estimate of the p quantile, $\hat{x}_p$, is usually the $p \cdot (N + 1)$–st order statistic. In [116] it is pointed out, however, that to avoid spurious skewness in $\hat{x}_p$, a correction must be made. Specifically, if the observations X_i are arranged in increasing order, the correction must be performed by identifying X_i with $\hat{x}_{q(i)}$ where

$$q(i) = \frac{2i - 1}{2N} \tag{5.82}$$

and then interpolating linearly to p from the two adjacent $q(i)$ values.

By estimating certain quantiles from the observations it is possible to estimate the stable parameters. The first quantile estimators were suggested by Fama and Roll [117] for *SαS* distributions with $1 \leq \alpha \leq 2$. They suggested estimating c by

$$\hat{c} = \frac{1}{1.654}[\hat{x}_{0.72} - \hat{x}_{0.28}] \tag{5.83}$$

where $\hat{x}_p$ ($p = 0.72, 0.28$) is the estimated p quantile of the *SαS* distribution. In [118] they showed this estimator to have an asymptotic bias of less than 4% and that it produced estimates that were asymptotically normally distributed.

The characteristic exponent, on the other hand, can be estimated from the tail behaviour of the distribution. Specifically, for some large f ($f = 0.95$ for example) Fama and Roll [117] proposed calculating first

$$\hat{z}_f = \frac{\hat{x}_f - \hat{x}_{1-f}}{2\hat{c}} \tag{5.84}$$

Given that the RV is *SαS* with characteristic exponent α and dispersion $\gamma = c^\alpha$, $\hat{z}_f$ is an estimate of the f quantile of the *standard SαS* distribution. We then obtain the estimate for the characteristic exponent from

$$\hat{\alpha} = G(f, \hat{z}_f) \tag{5.85}$$

where $G(f, \hat{z}_f)$ is some mapping function based on tables of standard *SαS* distributions, such as those in [117].

Fama–Roll's method is simple but suffers from a small asymptotic bias and is not asymptotically efficient. Also, α is restricted to $1 \leq \alpha \leq 2$. McCulloch [116] proposed a generalisation of Fama–Roll's method to provide consistent estimates for α and c. Specifically, the quantity

$$\hat{v}_\alpha = \frac{\hat{x}_{0.95} - \hat{x}_{0.05}}{\hat{x}_{0.75} - \hat{x}_{0.25}} \tag{5.86}$$

is independent of c. Thus, a consistent estimate $\hat{\alpha}$ can be found by searching tables, such as those in [116] with matched value of v_α. Furthermore, for fixed α,

$$v_c = \frac{\hat{x}_{0.75} - \hat{x}_{0.25}}{c} \tag{5.87}$$

depends only on α. Since $\hat{\alpha}$, $\hat{x}_{0.75}$, and $\hat{x}_{0.25}$ are all consistent estimators, the following is a consistent estimator of c:

$$\hat{c} = \frac{\hat{x}_{0.75} - \hat{x}_{0.25}}{v_c(\hat{\alpha})} \tag{5.88}$$

Here, $v_c(\hat{\alpha})$ can be found in tables for an estimate of the characteristic exponent $\hat{\alpha}$.

5.9.2 Characteristic function based techniques

It is known that the characteristic function in eq. (3.1) (page 22) uniquely defines the distribution of a random variable X. Conversely, it should be possible to estimate the stable parameters from the empirical characteristic function. Thus, the observations X_i may be used to estimate $\Phi_\alpha(\omega)$. The sample characteristic function is defined as

$$\hat{\Phi}_\alpha(\omega) = \frac{1}{N}\sum_{i=1}^{N} \exp(j\omega X_i) \tag{5.89}$$

where N is the sample size. The sample characteristic function $\hat{\Phi}_\alpha(\omega)$ is a stochastic process with the useful property that $0 < |\hat{\Phi}_\alpha(\omega)| \leq 1$. So, all moments of $\hat{\Phi}_\alpha(\omega)$ are finite.

Techniques based on the sample characteristic function were first developed by Press [119]; another method was developed by Paulson, Holcomb and Leitch [120]. Finally, the regression method was developed by Koutrouvelis [121, 122]. It has been shown in [123] that Koutrouvelis' regression method is better than the other two in terms of consistency, bias and efficiency.

Koutrouvelis' regression method is based on the following relation

$$\ln\left\{-\ln|\Phi_\alpha(\omega)|^2\right\} = \alpha\ln|\omega| + \ln 2c^\alpha \tag{5.90}$$

which is obtained by twice taking the natural logarithm of $|\Phi_\alpha(\omega)|^2$ in eq. (3.9). Therefore, the parameters α and c can be estimated from the linear regression

$$y_k = \alpha x_k + \mu + \epsilon_k \quad , \quad k = 1,\ 2,\ \ldots\ ,\ K \tag{5.91}$$

where

$$y_k = \ln\left(-\ln|\hat{\Phi}_\alpha(\omega_k)|^2\right), \quad x_k = \ln|\omega_k|, \quad \mu = \ln 2c^\alpha \tag{5.92}$$

The sequence $\{\epsilon_k\}$ denotes the error terms which are assumed to be i.i.d. with zero mean. ω_1, $\omega_2, \ldots, \omega_K$ are an appropriate set of real numbers. This regression can be solved with the least squares method. Let $\mathbf{y}$ and $\boldsymbol{\epsilon}$ be the vectors containing the samples y_k and ϵ_k, respectively. Define now

$$\mathbf{X} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_K & 1 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} \alpha \\ \mu \end{bmatrix} \tag{5.93}$$

Then eq. (5.91) can be written in matrix form

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \boldsymbol{\epsilon} \tag{5.94}$$

Then, the least squares estimate of the vector of unknowns $\mathbf{w}$ is

$$\hat{\mathbf{w}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \tag{5.95}$$

hence

$$\hat{\alpha} = \hat{\mathbf{w}}_1 \tag{5.96}$$

$$\hat{c} = \exp\left(\frac{\hat{\mathbf{w}}_2 - \ln 2}{\alpha}\right) \tag{5.97}$$

One disadvantage of this method is that the choice of ω_k depends on the value of c. This is because the scale parameter can render the characteristic function very flat and if ω_k are not chosen correctly, the estimates can be very unstable [58]. To overcome this, the data can be normalised, i.e.,

$$X \to \frac{X}{\hat{c}} \tag{5.98}$$

where $\hat{c}$ is the estimate of c from McCulloch's quantile method. The ω_k can then be selected

equi–spaced over the interval $[0.1,\ 1.0]$ [58].

5.9.3 Fractional lower order moments based techniques

The fractional lower order moments (FLOM's) were presented in section 3.6. Like the characteristic function, their values are unique for a given set of stable parameters. Hence their empirical approximation can be used for estimating the stable parameters.

The log FLOM method was proposed by Ma and Nikias [124]. If X is a real $S\alpha S$ RV, then its p–th order moment $\boldsymbol{E}\left\{|X|^p\right\}$ can be written as $\boldsymbol{E}\left\{\mathrm{e}^{p\ln|X|}\right\}$, where $-1 < p < \alpha$. The proof that finite FLOM's also exist for $p < 0$ is given in [124]. Thus, a new variable can be defined as $Y = \ln|X|$. It is shown that the expected value of Y is given by

$$\boldsymbol{E}\left\{Y\right\} = C_e\left(\frac{1}{\alpha} - 1\right) + \frac{1}{\alpha}\ln\gamma \tag{5.99}$$

where $C_e = 0.57721566\ldots$ is the Euler constant. Moreover, the variance of Y can be written as

$$\mathrm{Var}(Y) = \frac{\pi^2}{6}\left(\frac{1}{\alpha^2} + \frac{1}{2}\right) \tag{5.100}$$

But, it is possible to substitute the sample mean $\bar{Y}$ and the sample variance $\hat{\sigma}_Y^2$ of Y in equations (5.99) and (5.100). By solving eq. (5.100) we can obtain an estimate for α and substitute into eq. (5.99) for an estimate of γ.

The second method uses asymptotic extreme value theory to estimate α and a FLOM to estimate γ. This technique was developed by Tsihrintzis and Nikias [125, 126]. To estimate α the data is divided into L equi–lengthed blocks of length $K = N/L$. In each block the maximum and minimum values are found: $\overline{X}_l$ and $\underline{X}_l$, respectively ($l = 1, 2, \ldots, L$). Then let $\overline{x}_l = \ln\overline{X}_l$ and $\underline{x}_l = \ln\underline{X}_l$. If this is done for all l, define

$$\overline{s} = \sqrt{\frac{1}{L-1}\sum_{l=1}^{L}(\overline{x}_l - \overline{x})^2} \quad , \quad \overline{x} = \frac{1}{L}\sum_{l=1}^{L}\overline{x}_l \tag{5.101}$$

and

$$\underline{s} = \sqrt{\frac{1}{L-1}\sum_{l=1}^{L}(\underline{x}_l - \underline{x})^2} \quad , \quad \underline{x} = \frac{1}{L}\sum_{l=1}^{L}\underline{x}_l \tag{5.102}$$

Then, the estimate for the characteristic exponent is given by

$$\hat{\alpha} = \frac{\pi}{2\sqrt{6}}\left(\frac{1}{\overline{s}} + \frac{1}{\underline{s}}\right) \tag{5.103}$$

The dispersion is estimated using a FLOM based technique. Tsihrintzis noted that

$$\hat{\gamma} = \left(\frac{\frac{1}{N}\sum_{k=1}^{N}|X_k - \hat{\alpha}|^p}{C(p,\hat{\alpha})}\right)^{\frac{\hat{\alpha}}{p}} \tag{5.104}$$

where

$$C(p,\hat{\alpha}) = \frac{1}{\cos\left(\frac{\pi}{2}p\right)} \cdot \frac{\Gamma\left(1 - \frac{p}{\hat{\alpha}}\right)}{\Gamma(1-p)} \tag{5.105}$$

So, by choosing a suitable value for p, such that $p < \alpha$, an estimate for γ can be obtained.

5.9.4 Discussion

In this section we presented a variety of algorithms that have been developed in the literature for the estimation of stable parameters. These algorithms are based either on statistical quantiles, or on the sample characteristic function of the data, or on fractional lower order moments.

The quantile based techniques, although efficient in a statistical analysis environment, are not suitable for signal processing in a communications context. Firstly, they involve order statistics which have to be calculated on the full set of received data. This requires a storage memory and number of computations which grow with time.

The characteristic function based scheme (Koutrouvelis' algorithm), on the other hand, has been formulated as a linear regression problem, and its requirements in storage and computations do

not grow with time. This characteristic and the fact that the implementation of this algorithm is straightforward are highly desirable in signal processing for communications. Furthermore, its estimates are consistent and unbiased [123].

For example, fig. 5.17 depicts the evolution of the characteristic exponent estimate using Koutrouvelis' scheme. The experiment was repeated for a variety of values for α, from pure Gaussian ($\alpha = 2$) to highly impulsive ($\alpha = 0.25$), with a constant $c = 1.5$. Accordingly, fig. 5.18 shows the estimates for various values of c, with constant $\alpha = 1.5$. As the results show, the algorithm needs less than 1000 samples to converge to a value close to the true parameter, while its performance does not change with α. However, as mentioned in section 5.9.2, the choise of ω_k's depends on c, so a rough estimate of the process scale is required.

In terms of efficient implementation and simplicity, however, the log FLOM algorithm, proposed by Ma and Nikias, is superior. This is a purely recursive algorithm, with minimal computational complexity and fairly simple implementation. Its main disadvantage, though, is depicted in fig. 5.19: the convergence speed of the characteristic exponent estimate degrages for α close to 2. Nevertheless, the estimation of c (fig. 5.20) is more robust, even though its computation involves the estimate for α.

Consequently, Koutrouvelis' method offers superior performance, but at the cost of complexity. On the other hand, the log FLOM algorithm, although simple, suffers from poor performance when the noise is close to Gaussian. However, as it will be shown in the following chapter, the sensitivity of the optimum Bayesian equaliser described in chapter 4 to the estimate of the characteristic exponent is small enough to accommodate the inaccuracy of this algorithm's estimates. Hence, the adoption of the algorithm proposed by Ma and Nikias is considered adequate for estimating the stable parameters in a Bayesian equaliser. Therefore, this algorithm will be used in chapter 6 in order to evaluate the performance of the adaptive Bayesian equaliser in α–stable noise environments.

5.10 Conclusions

The training of the optimum Bayesian equaliser in an α–stable noise environment was addressed in this chapter. Recall that the parametric implementation of the Bayesian equaliser was discussed in section 4.3. In particular, fig. 4.8 (page 58) depicts all sets of parameters employed by this scheme. These include the channel noise–free vector states $\mathbf{c}_i$ with their associated signs

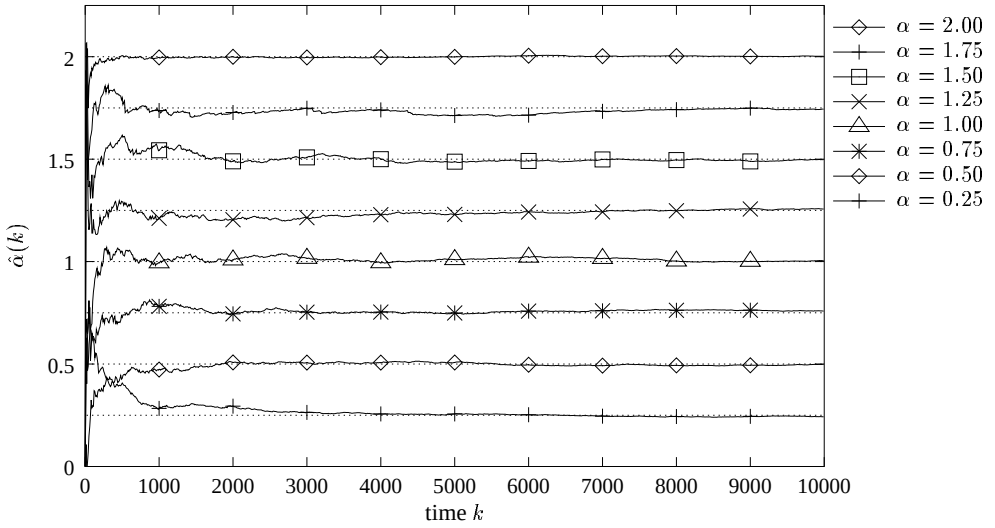


Figure 5.17: *Estimation of α by Koutrouvelis method.*

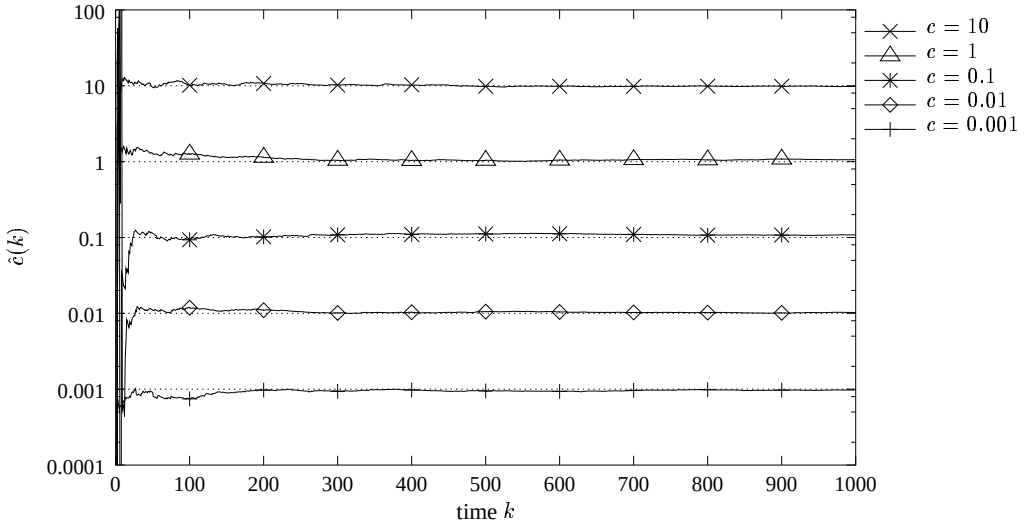


Figure 5.18: *Estimation of c by Koutrouvelis method.*

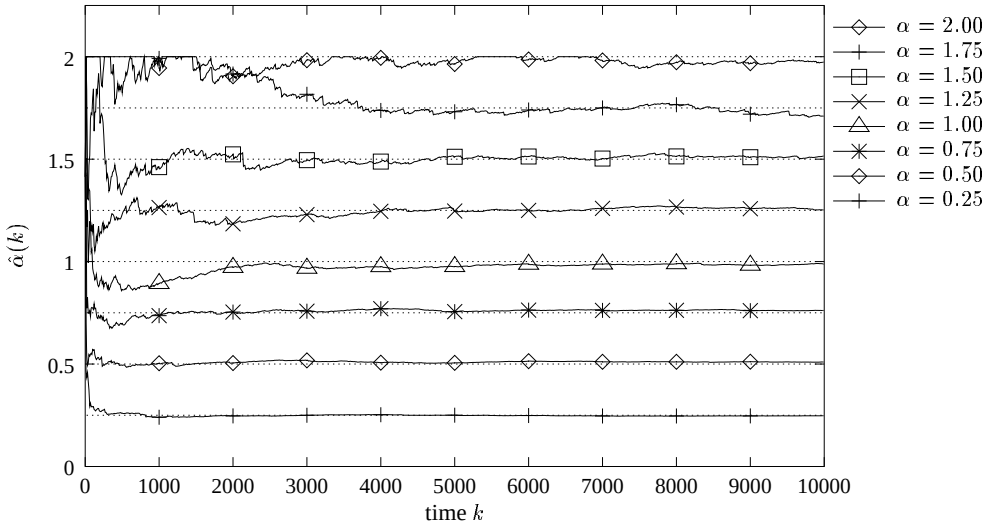


Figure 5.19: *Estimation of α by Ma–Nikias method.*

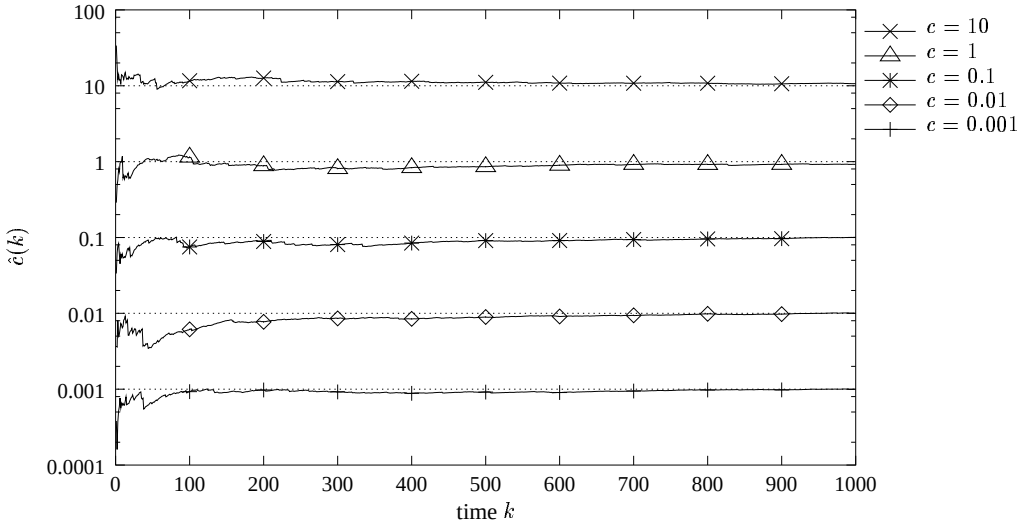


Figure 5.20: *Estimation of c by Ma–Nikias method.*

s_i, and the shape parameters of the M–dimensional α–stable multivariate distribution.

The estimation of the equaliser centers can be achieved by two principal techniques. In the indirect method the centers are inferred from an estimate of the channel impulse response, which is made by a traditional linear estimation algorithm. The direct method, on the other hand, applies a clustering algorithm on the received data to formulate the set of centers. Both schemes were studied, and the means for providing robust behaviour in non–Gaussian noise environments were investigated.

For the indirect method, linear estimation algorithms for non–Gaussian signals are needed, and a variety of such algorithms can be found in the literature. Most of them are reformulations of traditional techniques designed for Gaussian signals, which in general attempt to assign less importance to some *outlying* [2] observations. However, the criteria for the identification of outliers often involve *ad hoc* procedures.

The focal point for the first part of this chapter was formulating the channel estimation as an optimisation problem. Under this approach, we were able to show that a variety of algorithms can be expressed as derivations of the *sum of weighted least squares* cost function (eq. (5.29)), which assigns a specific weight (importance) to every observation. Moreover, with the appropriate choice of the weights, this cost function can implement a variety of optimisation criteria. Unfortunately, only the *minimum sum of squares* leads to analytically tractable and accurate solutions.

Nevertheless, it is possible to build a recursive process (the *recursive weighted least squares algorithm*, or RWLS) which approximates the optimum solution to any general optimisation criterion. The most important advantage of RWLS is its remarkable similarity to the familiar recursive least squares (RLS) scheme, which makes it an ideal choice for digital communications signal processing.

Although the maximum likelihood criterion is statistically optimum, it is computationally inefficient in an α–stable noise environment. This is because it requires the computation of the stable distribution and its derivative, which do not exist in closed form. However, as shown in [9], the *least p–norm* criterion is also meaningful for α–stable signals in the statistical sense, but most importantly turns out to be fairly easy to calculate.

[2] i.e., samples that contradict the Gaussian law.

The combination of RWLS with the least p–norm criterion gives rise to the *recursive least p–norm* (RLP) algorithm. Although, there is no convergence or stability analysis available, RLP demonstrates reliable and fast convergence behaviour in a non–Gaussian α–stable environment. Moreover, the performance of the mean squared estimation error for this algorithm is very close to that of the *iteratively re–weighted least squares* (IRLS) scheme, but at a considerably lower computational cost. Therefore, we propose this algorithm as the ideal alternative to the stochastic gradient based techniques.

For the estimation of the equaliser centers with the direct clustering scheme, our experiments suggested that the performance of the conventional algorithms degrades seriously in non–Gaussian environments. As a means for making these algorithms robust we propose introducing a *median filter* at the input to the clustering algorithm. Although effective, this method lacks theoretical justification and is not efficient in implementation. Alternatively, a least p–norm approach can be taken for the direct estimation of the centers. Specifically, we derived a least p–norm location estimation algorithm. This algorithm is actually a reduction of RLP , and is very efficient to implement. Moreover, its performance was shown to be robust and its tracking capability excellent.

Finally in this chapter, a comparative presentation of the existing algorithms for estimating the stable parameters was conducted. Although a variety of techniques have been developed in the literature, few of them enjoy efficiency, simplicity and low complexity at the same time. The best candidate for integration into the equalisation problem is the algorithm devised by Koutrouvelis [121, 122], but requires preprocessing of the data, as well as a large number of computations. By contrast, the algorithm proposed by Ma and Nikias [124] is particularly efficient in implementation, but suffers from poor performance when the noise is close to Gaussian. However, the Bayesian equaliser demonstrates a low sensitivity to the stable parameters, as will be shown in chapter 6. Therefore, the latter algorithm can be regarded as the ideal choice for estimating the stable parameters in a Bayesian equaliser.

Chapter 6
Performance evaluation of the adaptive Bayesian equaliser

For the problem of channel equalisation in a non–Gaussian stable noise environment, the optimum symbol–by–symbol detector was derived in chapter 4 in the form of the Bayesian (or MAP) equaliser. We presented experimental results suggesting that the optimum equaliser offers a considerable performance benefit compared to the traditionally designed equaliser (i.e., under the Gaussian assumption) for such impulsive noise environments. In these early experiments, however, the receiver had perfect knowledge of the channel and noise characteristics. In a realistic scenario the receiver would be required to estimate these characteristics from the actual received data.

Chapter 5 addressed the problem of estimating the channel and noise parameters in α–stable noise environments. For the former, our simulations showed that the proposed generalised versions of traditional channel estimation algorithms actually converge and produce satisfactory estimates of the channel. Moreover, for the estimation of the α–stable parameters a variety of techniques were presented, with the algorithm proposed by Ma and Nikias [124] being the main candidate for digital communication applications.

The question now rises naturally: how much of the promised performance gain can be practically attained by an adaptive system combining i) a non–Gaussian MAP equaliser, and ii) a set of algorithms for the estimation of the channel and noise parameters from the actual received data? Such a complete adaptive equaliser was shown in fig. 4.11 (page 65). This chapter attempts to provide useful insight on this question presenting the results from a variety of experiments performed in diverse environments. This chapter also discusses some practical issues regarding the implementation of such an adaptive equaliser. In particular, the fact that, in general, the α–stable pdf is not known in closed form makes certain calculations impossible. In this direction, we propose some useful approximations concerning certain practical aspects of the implementation of the Bayesian equaliser.

6.1 Experiments with correct noise parameters

In this section we present results from a set of simulation experiments, carried out in order to investigate the performance of the adaptive Bayesian DFE with correct noise parameters but estimated channel response. The transmitted data were organised in frames of 128 bits with the first 32 bits serving as training preamble. The bit rate is assumed to be 300 kbps. Both stationary and Rayleigh time–varying scenarios were simulated. For the latter, the taps of the non–stationary channel were correlated Rayleigh random variables multiplied by the appropriate tap root–mean–power (RMP).

The Rayleigh random variables were generated (as in [101]) by the envelope of a correlated complex–valued Gaussian random variable. The correlation of the Gaussian sequences was introduced by a second order low pass Butterworth filter. The bandwidth of this filter is of the order of the maximum Doppler frequency. In the simulations a cut–off frequency of 100 Hz is used. The output of the low pass filter is then scaled so that the variance of the correlated Gaussian random variable is 1. For the non–stationary channel scenario, the signal–to–noise ratio is defined as in [101], specifically

$$\mathrm{SNR}_{\mathrm{obs}} = \frac{\lim\limits_{k\to\infty} \dfrac{1}{k} \displaystyle\sum_{i=0}^{k-1} \boldsymbol{E}_x \left\{ |r_{\mathrm{L}}(k)|^2 \right\}}{\boldsymbol{E}_n \left\{ |\hat{n}(k)|^2 \right\}} \tag{6.1}$$

where $\boldsymbol{E}_p \{\cdot\}$ denotes the expectation operator with respect to the random process p. Recall that $r_{\mathrm{L}}(k)$ is the limited received signal and $\hat{n}(k)$ is the estimate of the noise sequence (see fig. 4.11, page 65).

For the channel estimator two algorithms were used: the modified least p–norm (MLMP) and the recursive least p–norm (RLP). The adaptation of the channel estimate taps takes place in both the training period and the data transmission period. For the former the known training sequence is used and for the latter the actual decisions of the equaliser are fed into the algorithm (decision–directed adaptation). The order of the forward section of the equaliser was set equal to the actual length of the channel ($M = 3$). For the order of the feed–back section (D) and the decision lag (d) the guidelines in [80, 127] were used ($d = N - 1 = 2$, $M = d + 1 = N = 3$, and $D = N + M - d - 2 = N - 1 = 2$).

The following abbreviations will be used for the figures in this chapter: “Trad. DFE” denotes

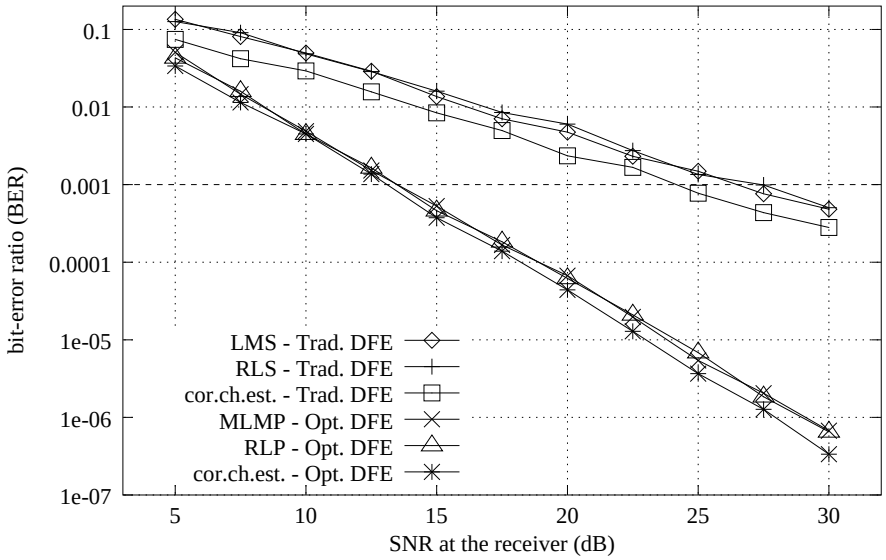


Figure 6.1: *Performance of the adaptive Bayesian DFE for a stationary channel with 3 taps and* $\alpha = 1$ *(*$M = 3, D = 2, d = 2, G = 4$*).*

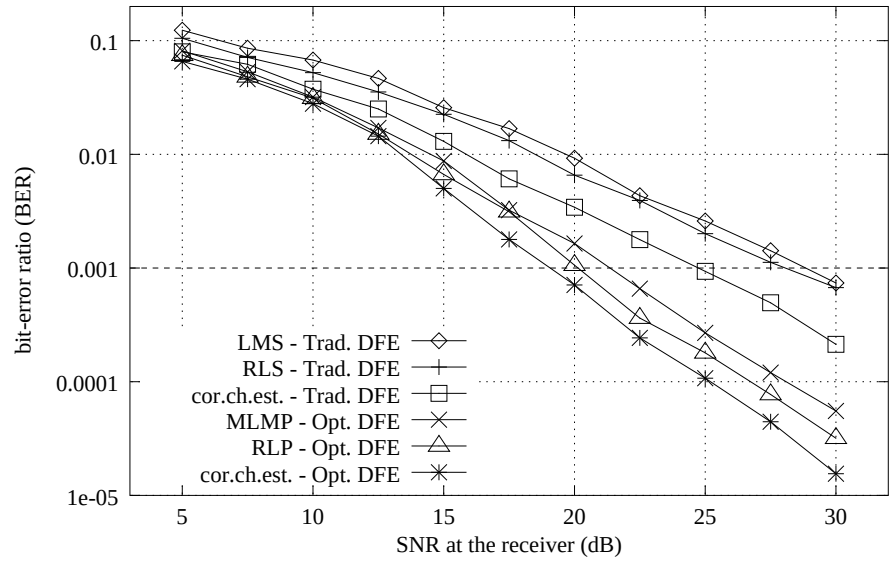
a traditionally designed Bayesian decision feedback equaliser, that is assuming the noise is Gaussian. "Opt. DFE" refers to an optimally designed Bayesian DFE, assuming that the noise is α–stable. "cor. ch. est." refers to correct channel estimation, that is the equaliser has perfect knowledge of the channel.

Stationary channel: The stationary channel consists of 3 taps

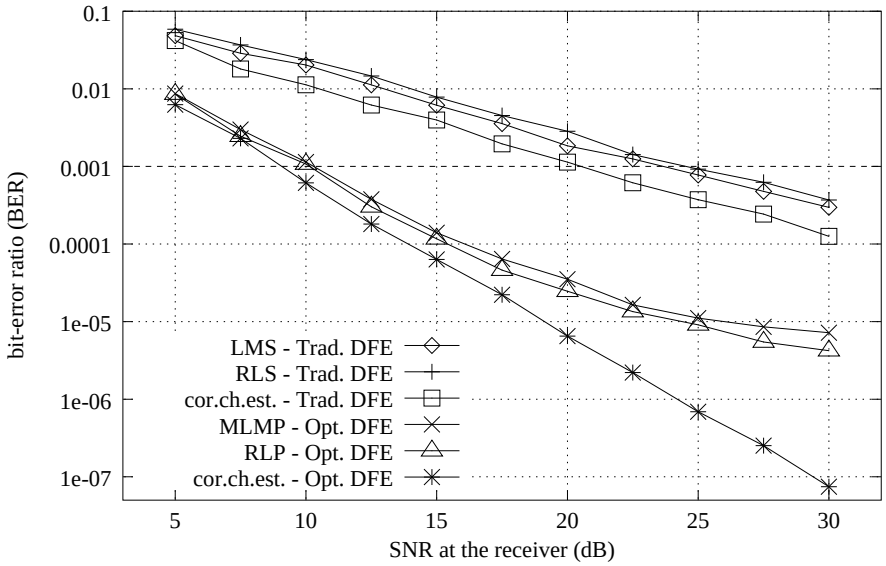
$$\mathbf{h} = [0.3482 \quad 0.8704 \quad 0.3482]^{\mathrm{T}} \tag{6.2}$$

and the noise characteristic exponent is $\alpha = 1$. The limiting level of the saturation device was chosen $G = 4$. Figure 6.1 depicts the performance of both optimum and traditional adaptive Bayesian DFE's. The performance of the equalisers with a perfect channel estimation is given as well.

The optimum adaptive DFE has a performance which is very close to the optimal and it seems that both MLMP and RLP perform equally well in a stationary environment. On the other hand, the traditional adaptive DFE suffers a significant performance degradation in this highly impulsive noise environment. For example, at a BER target of 10^{-3} this degradation is 13.34 dB (for RLP and RLS channel estimators, respectively) and 12.99 dB (for MLMP and LMS channel estimators, respectively).



(a) $\alpha = 1.5$



(b) $\alpha = 1$

Figure 6.2: *Performance of the adaptive Bayesian DFE for a Rayleigh fading channel with 3 taps: a)* $\alpha = 1.5$*, and b)* $\alpha = 1$ *(*$M = 3, D = 2, d = 2$ *and* $G = 6$*).*

Rayleigh fading channel The Rayleigh fading channel consists of 3 taps with RMP's

$$\mathbf{h} = [0.3482 \quad 0.8704 \quad 0.3482]^{\mathrm{T}} \tag{6.3}$$

The performance of both the optimum and traditional adaptive Bayesian DFE was recorded in a noise environment with characteristic exponent $\alpha = 1.5$ and $\alpha = 1$ (fig. 6.2(a) and fig. 6.2(b), respectively). The performance of the equalisers with a perfect channel estimation is shown as well. As expected, there is a definite performance loss of the adaptive DFE in comparison to the non-stationary scenario, due to the limited tracking ability of the channel estimation algorithms and the fading characteristics of the channel. However, the performance advantage of the optimum adaptive Bayesian DFE is still significant, compared to the bit–error ratio of the corresponding traditional DFE (i.e. designed under the Gaussian assumption). This benefit for $\alpha = 1.5$ is 7.93 dB at 10^{-3} target BER with RLP and 7.48 dB with MLMP. For $\alpha = 1$ the performance gain is 14.39 dB for RLP and 13.38 dB for MLMP at a BER of 10^{-3}.

The tracking performance of RLP in this fast changing environment is marginally better than MLMP, especially for the highly impulsive noise environment $\alpha = 1$ (fig. 6.2(b)). This is exactly the opposite situation to the Gaussian noise environment [101], where the stochastic gradient algorithm achieves better tracking of the channel than the least squares approach. This dissimilarity should be attributed to the noise statistics and the actual cost function of the channel estimation algorithms.

The principal consequence of a non-quadratic cost function, though, is a noticeable deterioration of the tracking ability of the adaptive algorithms. Recall from section 5.7 that in a highly impulsive noise environment, the weighting sequence $w(i)$ (eq. (5.36)) suppresses the samples with large estimation error, because they are likely to be the result of noise impulses. In particular, the more α approaches 1, the more large residuals are suppressed. But, when the channel is non-stationary, large residuals often arise as a result of the discrepancy between the channel estimate and the actual channel impulse response. The suppression of these residuals effectively decelerates the adaptation of the channel estimation algorithms.

6.2 Sensitivity to the noise parameters estimates

Before introducing the noise parameters estimation algorithm into the adaptive equaliser, it would be useful to investigate the sensitivity of the Bayesian detector to these noise parameters

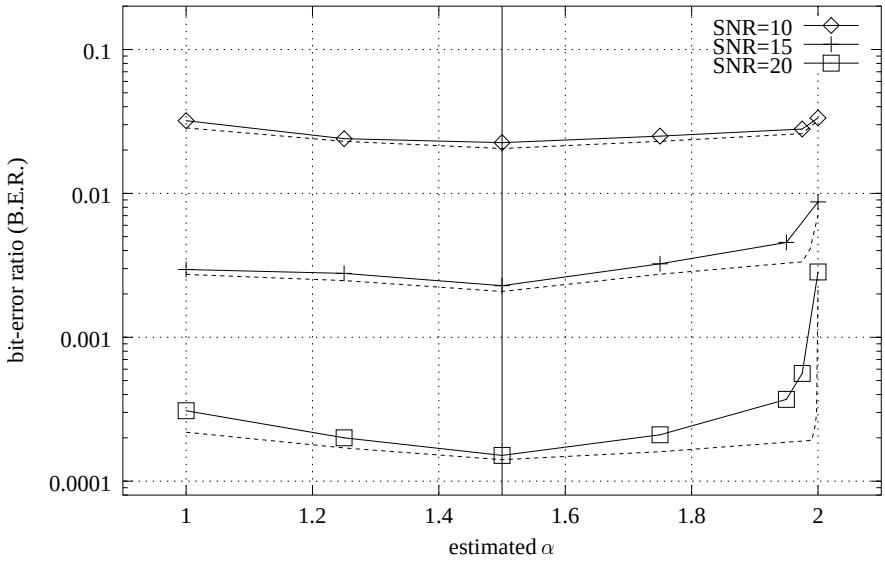


Figure 6.3: *Robustness of the adaptive (RLP) Bayesian DFE equaliser (solid lines) with respect to the estimated characteristic exponent α for actual $\alpha = 1.5$. The dashed lines correspond to perfect channel knowledge.*

(i.e. α and γ). For this purpose the same constellation as in section 6.1 was used. Throughout the experiment the true parameters α and γ were fixed. Then, the performance of the Bayesian equaliser was recorded for a range of estimated values for these parameters.

Figure 6.3 shows the effect of the characteristic exponent estimate on the performance of the MAP equaliser. Here, the true value of this parameter is $\alpha = 1.5$. The solid lines correspond to a Bayesian equaliser operating with an estimate for the channel response, while the dashed lines depict the performance of the same equaliser with perfect knowledge of the channel. As the figure shows, the MAP section of the receiver is quite robust with respect to the estimate of the characteristic exponent, as long as $\alpha < 2$ (fig. 6.3 – dashed lines). On the other hand, as shown in [9], all lower order moments of a *S*α*S* random variable are equivalent, i.e., the p–th and q–th order moments differ by a constant factor independent of the *S*α*S* random variable for all $0 < p, q < \alpha$. This important result implies, in general, that an arbitrary p such that $0 < p < \alpha$ for the cost function $\mathcal{J}_{LP}$ in eq. (5.58) (page 96) would be equivalent to any other such p. That explains why the performance of the adaptive Bayesian DFE (fig. 6.3 – solid lines) deteriorates more severely when the characteristic exponent estimate is over–estimated rather than under–estimated.

However, in practice, the finite dynamic range receiver model implies that the received signal has finite variance. Consequently, even when the estimate for α is greater than its true value, the

channel estimation algorithm is not subject to infinite variance error signals and the convergence of the algorithm deteriorates less severely than the theory predicts. On the other hand, the convergence of the channel estimator deteriorates dramatically for $p \geq \alpha$ without the limiter at the front end of the receiver.

It has also been demonstrated in [101] that the Bayesian equaliser (operating in a Gaussian environment) is sufficiently robust against the estimation of the noise variance. Our experiments suggest that the same holds for non-Gaussian stable noise, as well.

6.3 Experiments with estimated noise parameters

For this set of experiments only the non–stationary channel scenario was studied. The taps of the channel are modelled again as Rayleigh RV's with unit variance, multiplied by the appropriate tap root–mean–power (RMP). Recall that a Rayleigh RV is the envelope of a complex coloured Gaussian RV.

The method for the generation of coloured Gaussian RV's used in section 6.1, although simple, suffers from some important drawbacks. First, it does not provide the scope for flexibility during the simulations, because changing the maximum Doppler frequency implies redesigning the low–pass filter, a procedure clearly not parametric. On the other hand, the cut–off frequency of the filter should be several orders of magnitude lower than the operating symbol rate. This makes the design of the filter an even more challenging task. Finally, this method cannot provide a profile for the psd of the Gaussian RV that is close to either of the proposed models in the literature [128].

The approach taken hereinafter for the generation of the real and imaginary parts of the complex coloured Gaussian RV's is deterministic and is based on Rice's sum of sinusoids [129, 130]. The statistical properties of this scheme are extracted by Pätzold et. al. in [131]. In this method, a real coloured Gaussian random process $m(t)$ is approximated by a finite sum of weighted sinusoids of the form

$$\hat{m}(t) = \sum_{n=1}^{N} c_n \ \cos(2\pi f_n t + \theta_n) \tag{6.4}$$

where the quantities c_n, f_n, and θ_n are called Doppler coefficients, discrete Doppler frequencies, and Doppler phases, respectively, and N denotes the number of sinusoids.

For the shape of the Doppler psd $S_{mm}(f)$ of the complex Gaussian noise process $m(t)$ we adopt the often–assumed Jakes psd [128] for mobile fading channel models, given by

$$S_{mm}(f) = \begin{cases} \dfrac{2\sigma_0^2}{\pi f_{\max}\sqrt{1-(f/f_{\max})}} & , \quad |f| \leq f_{\max} \\ 0 & , \quad |f| > f_{\max} \end{cases} \tag{6.5}$$

Pätzold et. al. showed in [131] that satisfactory approximation of the true statistical behaviour of a coloured Gaussian random process can be obtained (in terms of both amplitude probability density function and higher order statistical properties such as auto–correlation function, average duration of fades, level–crossing rate and pdf of the fading intervals) by the following choices for the model parameters

$$c_n = \sigma_0\sqrt{2/N} \quad , \quad \text{if } N \geq 7 \tag{6.6}$$

$$f_n = f_{\max}\sin\left[\frac{\pi}{2N}\left(n-\frac{1}{2}\right)\right] \tag{6.7}$$

$$\theta_n = 2\pi\frac{p_n}{N+1} \tag{6.8}$$

where $(p_1,\ p_2,\ \ldots\ ,\ p_N)$ is one of the $N!$ different permutations of the numbers $1, 2, \ldots, N$.

The key difference for this section's experiments compared with those in section 6.1, however, is that the Bayesian equaliser operates with estimates for the noise parameters. These estimates are produced directly from the received data. Specifically, the method developed by Ma and Nikias (presented in section 5.9.3, page 120) is applied on the noise estimate sequence $\{\hat{n}(k)\}$ (see fig. 4.11, page 65) in order to obtain estimates for the noise characteristic exponent α and dispersion γ. For the implementation of this algorithm, the sample mean and variance of the transformed variable $\ln|\hat{n}(k)|$ need to be calculated. For this, a simple adaptive scheme with exponential weighting has been chosen. The complete algorithm can be summarised as follows

Algorithm 6.1 (Adaptive noise parameters estimation)

Initialisation: $\mu(0) = -3.2625 \quad \sigma^2(0) = 1.5535$

$\hat{\alpha}(0) = 1.5 \quad \hat{\gamma}(0) = 0.01$

$P(0) = 1$

Basic recursion: $\forall\ k = 1, 2, \ldots$

1. weighting factor

$$P(k) = \frac{P(k-1)}{\lambda + P(k-1)}$$

2. sample mean

$$\mu(k) = \mu(k-1) + \left(\ln|\hat{n}(k)| - \mu(k-1)\right) P(k)$$

3. sample variance

$$\sigma^2(k) = \sigma^2(k-1) + \left(\ln^2|\hat{n}(k)| - \sigma^2(k)\right) P(k)$$

4. characteristic exponent estimate

$$\hat{\alpha}(k) = \sqrt{\frac{1}{\left|\frac{6\sigma^2(k)}{\pi^2} - \frac{1}{2}\right|}}$$

5. dispersion estimate

$$\hat{\gamma}(k) = \exp\left\{\hat{\alpha}(k)\left[\mu(k) - C_e\left(\frac{1}{\hat{\alpha}(k)} - 1\right)\right]\right\}$$

□

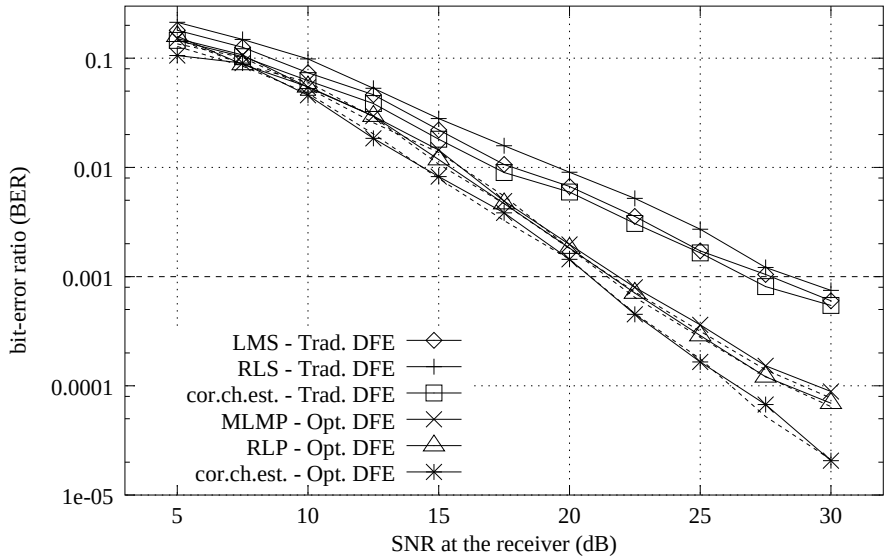
where $\lambda \in (0,\ 1]$ is an exponential weighting factor. The initial values for μ and σ^2 are chosen so that they correspond to the initial values of the noise estimates $\hat{\alpha} = 1.5$ and $\hat{\gamma} = 0.01$.

Two channel models were used, with tap RMP's

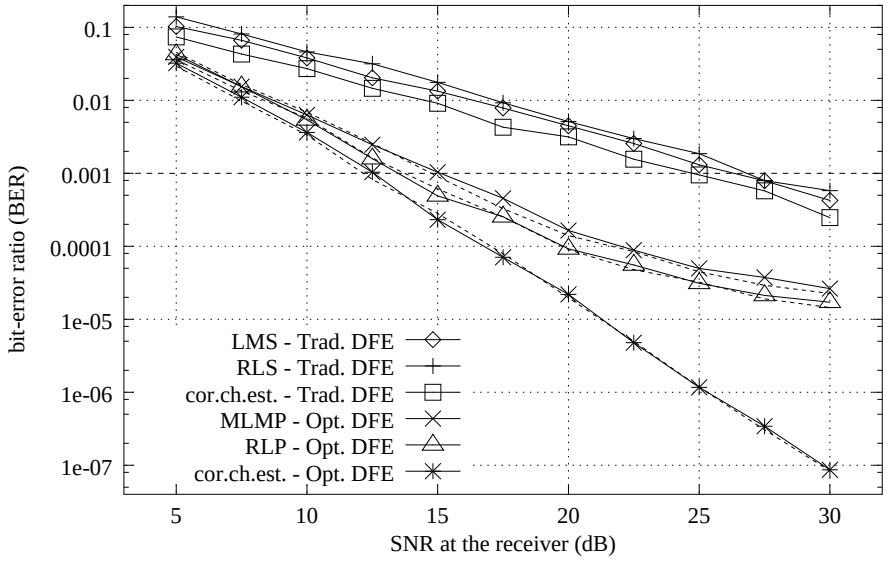
$$H_1(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2} \quad (6.9)$$

$$H_2(z) = 0.3333 + 0.6667z^{-1} + z^{-2} \quad (6.10)$$

The data have been organised in 128 bits frames, with 32 bits training preamble. Two different noise environments were simulated with characteristic exponents $\alpha = 1.5$ and $\alpha = 1$. The performance of the optimum Bayesian DFE, as well as the traditional Gaussian DFE was studied. For the first, two channel estimation algorithms were used: MLMP and RLP. For the latter these algorithms were LMS and RLS. The parameters for MLMP and LMS were set to $\mu_0 = 0.02$,

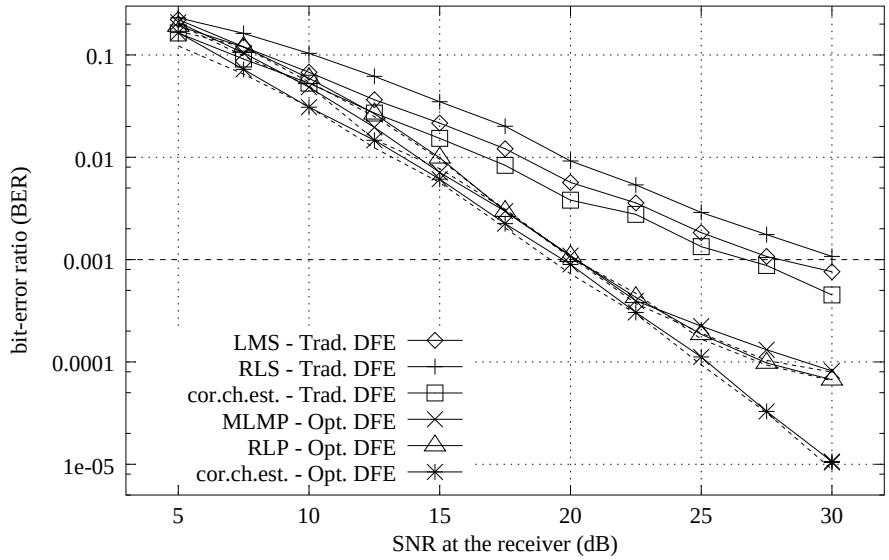


(a) $\alpha = 1.5$

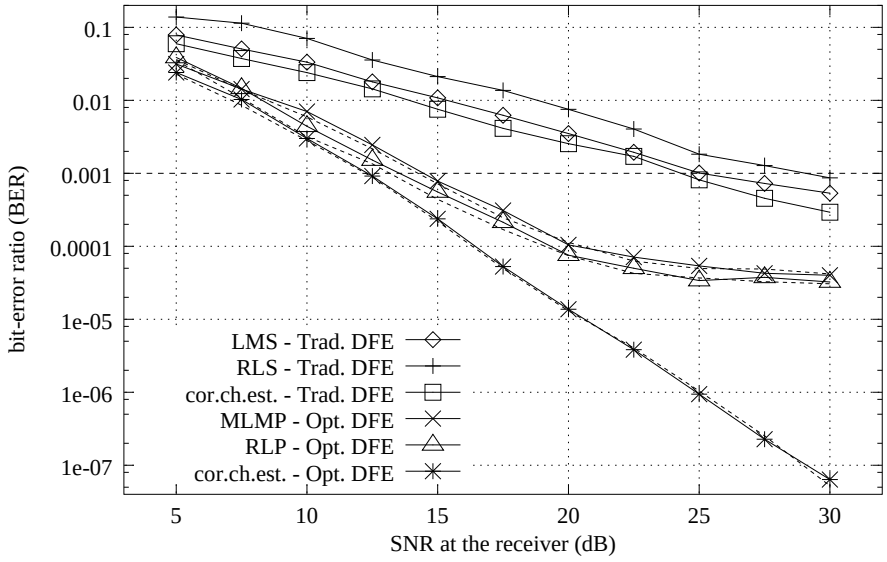


(b) $\alpha = 1.0$

Figure 6.4: *Performance of the adaptive Bayesian DFE with noise parameters estimates (solid lines) and actual noise parameters (dashed lines) for channel $H_1(z)$: a) $\alpha = 1.5$, and b) $\alpha = 1$ ($M = 3$, $D = 2$, $d = 2$ and $G = 6$).*



(a) $\alpha = 1.5$



(b) $\alpha = 1.0$

Figure 6.5: *Performance of the adaptive Bayesian DFE with noise parameters estimates (solid lines) and actual noise parameters (dashed lines) for channel $H_2(z)$: a) $\alpha = 1.5$, and b) $\alpha = 1$ ($M = 3$, $D = 2$, $d = 2$ and $G = 6$).*

$c = 44$, $\nu = 0.9$ and $\omega = 0.1$. For RLP and RLS we set $\delta = 0.05$, $\lambda = 0.875$ and $\omega = 0.05$. In all cases a symbol rate of $f_s = 300$ kbps was assumed and a maximum Doppler frequency $f_{\text{max}} = 100$ Hz.

Figures 6.4 (for channel $H_1(z)$) and 6.5 (for channel $H_2(z)$) depict the bit-error ratio performance of all algorithms for a range of SNR's at the receiver. For comparison, the performance of the equalisers with correct estimates of the noise parameters is given (dashed lines). These results indicate that the utilisation of estimates rather than the actual values for the noise parameters does not practically compromise the performance of the equalisers. Indeed, the plots are almost identical with those in fig. 6.2, suggesting that the dominant factor affecting the performance of the adaptive equaliser is the design of the Bayesian (MAP) detector and the channel estimation algorithms.

This result becomes more important if one takes into account that the algorithm employed for the estimation of the noise parameters does not offer the best attainable performance (especially in terms of the characteristic exponent - see fig. 5.19, page 124). Consequently, when the stable noise is non–Gaussian (i.e. $\alpha < 2$), only a rough estimate of the true characteristic exponent can provide near optimum performance. Besides, this is exactly the message drawn from fig. 6.3.

The above experimental setup was also used in order to explore the effect of the maximum Doppler frequency on the performance of the equalisers. The SNR at the receiver was set to 15 dB, the limiting level of the saturation device was chosen $G = 10$, the noise characteristic exponent $\alpha = 1$, and the channel was $H_1(z)$. The experiments were performed for a range of Doppler frequencies from 1 to 1000 Hz, and a symbol rate $f_{\text{s}} = 300$ kbps. For the MLMP algorithm the parameters were $\mu = 0.01$, $c = 89$, $\nu = 0.9$, and $\omega = 0.1$. For RLP we chose $\delta = 0.05$, $\lambda = .92$, and $\omega = 0.05$.

The results of this experiment are depicted in fig. 6.6. For this experiment the *tracking* parameters μ and λ were set so that both algorithms (MLMP and RLP) had the same effective averaging window size. It should be noted that for a given Doppler frequency, the variability of the channel is inversely proportional to the symbol rate. The performance benefit of the Bayesian DFE using RLP as a channel estimator is approximately 1.5 dB compared to MLMP for a wide range of channel variability.

This result supports the picture drawn from previous experiments and can be summarised as

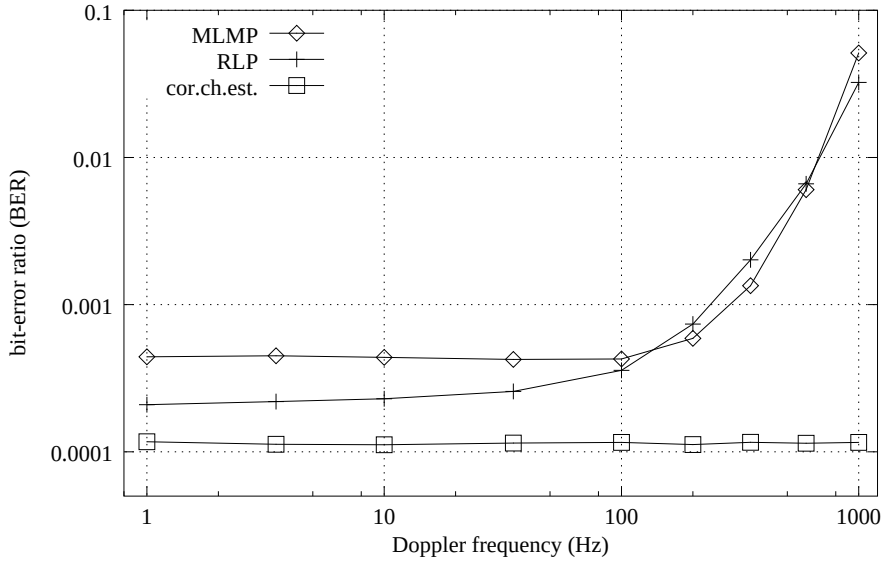


Figure 6.6: *The effect of the maximum Doppler frequency on the performance of the adaptive Bayesian equaliser for symbol rate* $f_s = 300$ *kbps and channel* $H_1(z)$.

follows: For a Gaussian environment the use of stochastic gradient algorithms in the channel estimator provides better performance for the Bayesian DFE compared to the least–squares technique. However, as the noise becomes more impulsive, the situation is reversed: for $\alpha = 1.5$ the performance of both schemes is indistinguishably close, while for $\alpha = 1$ the least–squares based techinque provides better performance. These conclusions are more thoroughly pictured in fig. 6.7 for two different values of SNR and channel $H_1(z)$. The figure shows the bit–error ratio of the Bayesian DFE with respect to the characteristic exponent for both channel estimation algorithms.

6.4 Practical approximations for stable distributions

Unfortunately, the performance benefit of the proposed equaliser comes at the expense of high computational load. Recall from chapter 3 that closed form α–stable densities only exist for $\alpha = 2$ (Gaussian) and $\alpha = 1$ (Cauchy). In all other cases, numerical approximation of the stable distribution is required, making the use of stable distributions in real time systems unaffordable. However, in MAP applications, only the actual shape of the decision boundary is important for the performance of the equaliser, which means that an approximation to the stable density may be used. In this section, we propose a linear interpolation between the Gaussian and the Cauchy

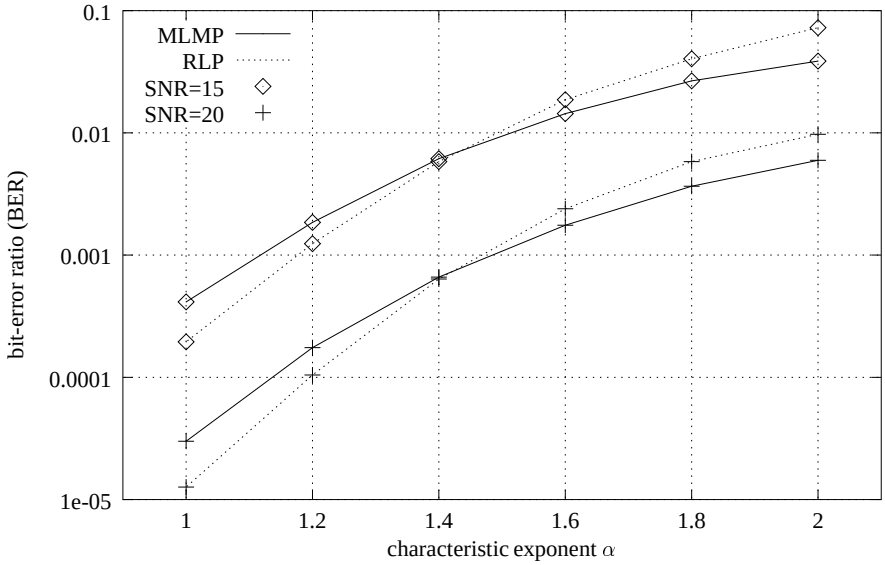


Figure 6.7: *Performance of the Bayesian DFE against the noise characteristic exponent with MLMP (solid lines) and RLP (dashed lines) for channel estimation.*

distribution as an approximation to the stable density for $1 < \alpha < 2$, i.e.

$$\hat{f}_\alpha(s) = (1-p)\ f_1(s) + p\ f_2(s) \tag{6.11}$$

where p is an increasingly monotonic function of α ($0 \leq p \leq 1$ for $1 \leq \alpha \leq 2$). $f_2(x)$ and $f_1(x)$ are the Gaussian and Cauchy distributions, respectively [1].

Assuming that the dynamic range of the receiver can accommodate all scalar centers without distortion ($G > |\bar{c}_i|,\ \forall i$), it is only required that the symmetric noise pdf is approximated within the range $[0, 2G]$. Furthermore, the shape of the noise distribution close to the origin does not affect the optimum decision boundary. Therefore, the approximation range can further be reduced to $[H, 2G]$, where $0 < H < 2G$. The optimum value for p with respect to α can then be derived by a least squares optimisation of the form

$$p_{\text{opt}}(\alpha) = \arg \min_{0<p<1} J_{\text{LS}} \tag{6.12}$$

where

$$J_{\text{LS}} = \int_H^{2G} \left| f_\alpha(s) - (1-p) f_1(s) - p f_2(s) \right|^2 ds \tag{6.13}$$

[1] A RV generated by time multiplexing a Gaussian RV with probability p and a Cauchy RV with probability $(1-p)$, actually has a pdf given by eq. (6.11).

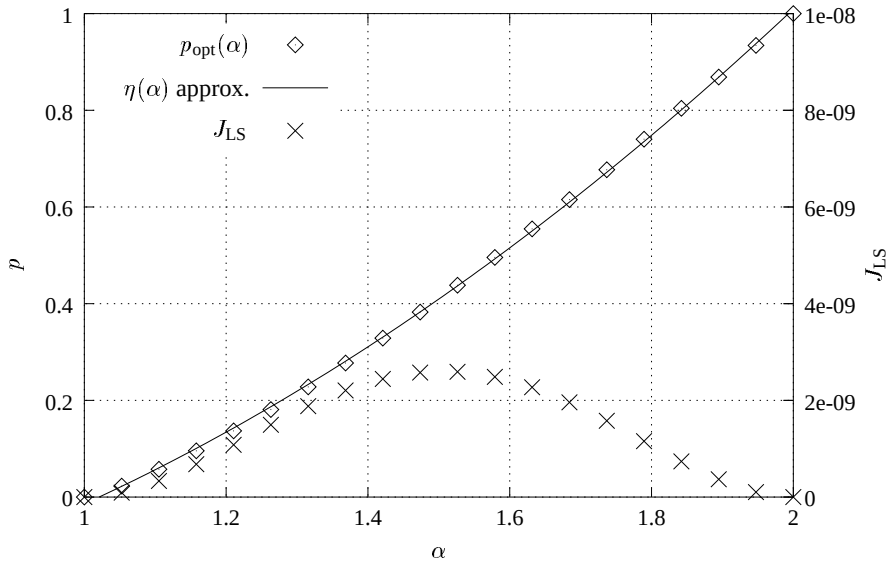


Figure 6.8: *The true and approximated p_{opt} as a function of α. The least squares estimation error J_{LS} is also depicted.*

is the least squares cost function.

We have numerically solved eq. (6.12) for a number of values of α with $G = 4$, $H = 2$, and $\gamma = 0.01$. The resulting set of optimum values for p is depicted in fig. 6.8 along with the minimum achieved least squares error J_{LS}. It would be desirable, however, to approximate $p_{\text{opt}}(\alpha)$ with a more simple formula. We can, for example, apply second degree polynomial fitting on the set of optimum values for p obtained from eq. (6.12) to produce a relation $p = \eta(\alpha)$. This relation has been found to be (see fig. 6.8)

$$\eta(\alpha) = 0.3521\ \alpha^2 - 0.0329\ \alpha - 0.3333 \tag{6.14}$$

Figure 6.9 shows the actual and the approximated pdf for $\alpha = 1.5$, where the fit is poorest, i.e $\min_{0<p<1} J_{\text{LS}}$ is maximum. However, it should be emphasised that the approximation of eq. (6.11) is only meaningful within the optimisation range $[-2G,\ 2G]$; outside this range, the approximation error can be very large. Furthermore, as fig. 6.10 shows, the approximated pdf produces a decision boundary which preserves the features of the optimum boundary. Fig. 6.11 depicts the BER performance of the Bayesian DFE in a stationary channel with 3 taps with the approximated pdf (solid lines) and the true pdf (dashed lines). These results suggest that the approximation of eq. (6.11) results in a performance loss of less than 1 dB for $\alpha = 1.5$. For $\alpha = 1.95$ the performance loss is indistinguishable.

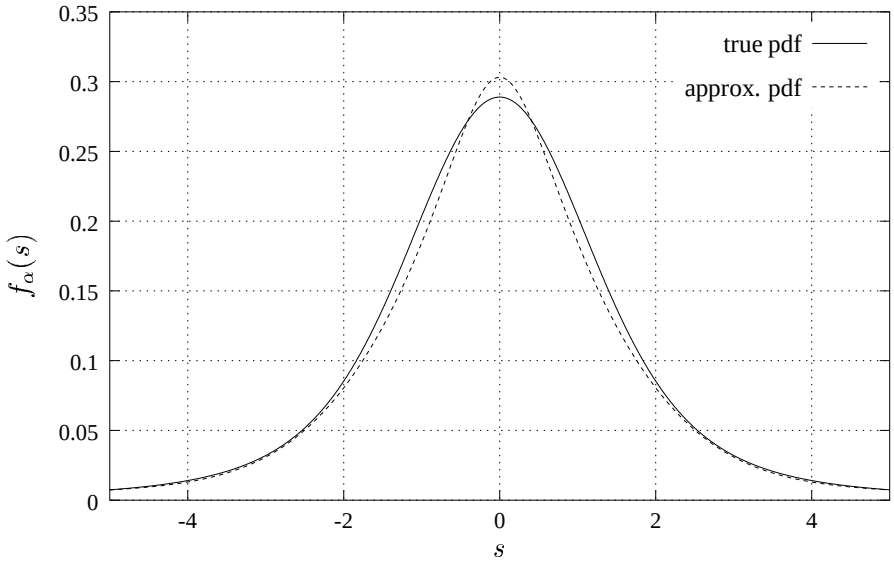


Figure 6.9: *The true and approximated α–stable pdf for $\alpha = 1.5$ ($\gamma = 1$).*

Interestingly, eq. (6.11) suggests an alternative model for signals exhibiting impulsive behaviour. As footnote 1 (page 140) explained, this is actually a mixture of a Gaussian and a Cauchy process, time multiplexed with probabilities p and $(1 - p)$, respectively. Therefore, the parameter p controls the heaviness of the tails in the resulting pdf. A value of p close to 1 defines a near–Gaussian model, while as p moves towards 0 the resulting probability density becomes more impulsive. Middleton [8], based on physical grounds, maintains that the impulsive noise experienced in nature has always a Gaussian component, although contributing in a different manner than eq. (6.11) suggests. However, if eq. (6.11) is regarded as more appropriate for modelling impulsive signals, the results presented hereto show that, reciprocally, the family of α–stable distributions can approximate this model to a satisfactory degree. In that manner, the algorithms for α–stable parameters estimation (described in chapter 5) can be utilised, since the estimation of the parameters of eq. (6.11) is a difficult theoretical problem.

In section 4.5, we analytically derived the variance of the noise estimate $\hat{n}$ (eq. (4.70)) for $\alpha = 1$ and $\alpha = 2$. In all other cases, it is only possible to measure this variance experimentally. Our experiments, however, suggest that as α moves from 1 to 2, the variance v_α of the noise estimate $\hat{n}$ moves in a linear way (in the $\log$ domain) from v_1 to v_2 (the calculated variances for $\alpha = 1$ and $\alpha = 2$, respectively). Therefore, a reasonable approximation should be

$$\hat{v}_{\hat{n}}(\alpha, \gamma, G) = v_{\hat{n}}(1, \gamma, G)^{2-\alpha}\, v_{\hat{n}}(2, \gamma, G)^{\alpha-1} \tag{6.15}$$

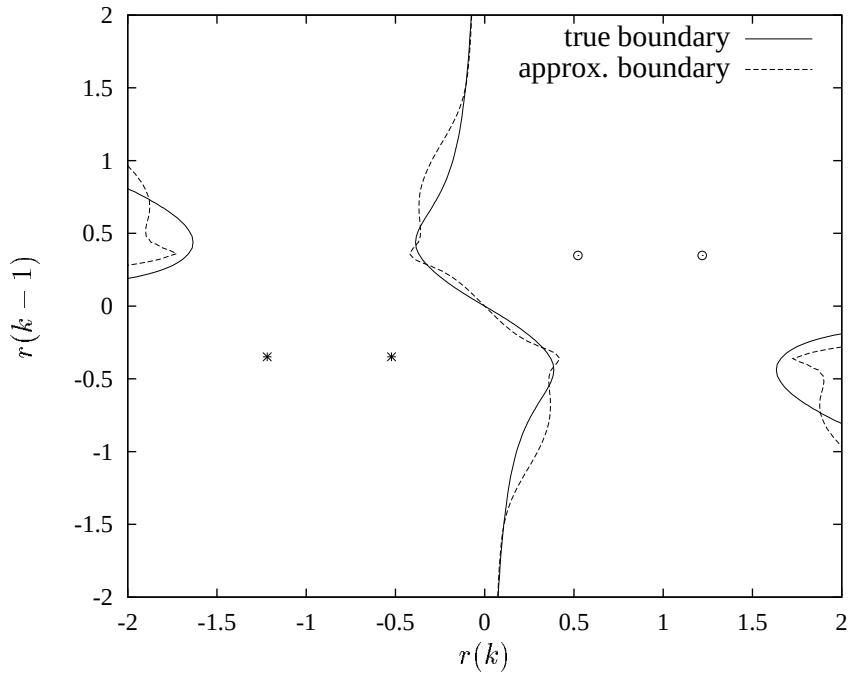


Figure 6.10: *The true and approximated decision boundary of the Bayesian DFE with channel* $H_1(z)$ *for* $\alpha = 1.5$ *(*$M = 2, D = 2, d = 1$*).*

Figure 6.12 depicts the experimental (true) $v_{\hat{n}}$ and approximated $\hat{v}_{\hat{n}}$ with respect to the noise dispersion γ for different values of α and $G = 4$. Figure 6.13, on the other hand, shows $v_{\hat{n}}$ and $\hat{v}_{\hat{n}}$ as a function of the limiting level G for different values of α. These graphs show that the approximation of eq. (6.15) is sufficiently satisfactory for a wide range of the limiting level G and noise dispersion γ.

Moreover, eq. (6.15) implies that given a *good* estimate for γ it is possible to obtain a fairly good estimate for α from the variance of the noise estimate $\hat{n}$ (and vice versa).

In a similar way to eq. (6.15), we can obtain a good approximation of the appropriate dispersion γ for a given $\mathrm{SNR}_{\mathrm{rcv}}$ when α is not equal to 1 or 2. More precisely, this approximation is

$$\hat{\gamma}_\alpha = \gamma_1^{2-\alpha}\, \gamma_2^{\alpha-1} \tag{6.16}$$

where γ_1 and γ_2 are the solutions of equation

$$v_{\hat{n}}(\alpha, \gamma, G) = \frac{v_y}{10^{\mathrm{SNR}_{\mathrm{rcv}}/10}} \tag{6.17}$$

with respect to γ for $\alpha = 1$ and $\alpha = 2$, respectively. Figure 6.14 shows that this estimate is

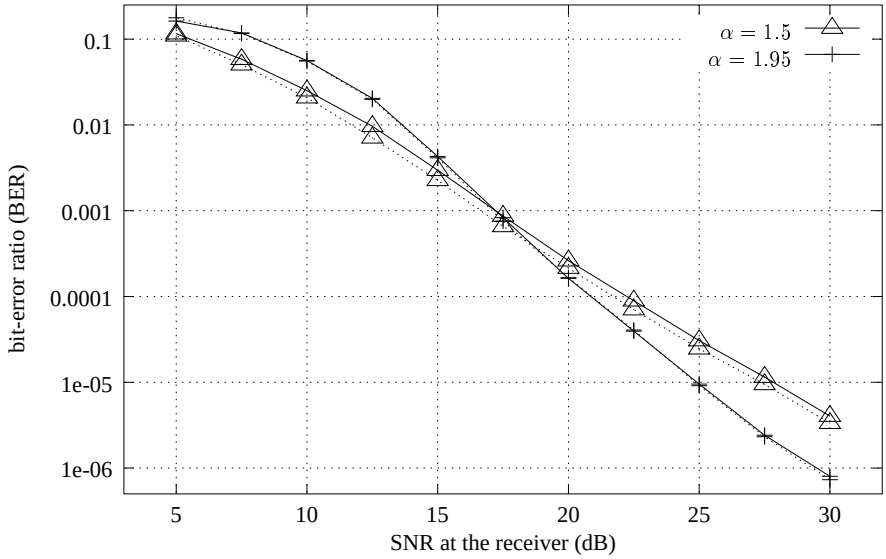


Figure 6.11: *Performance of the Bayesian DFE with true (solid lines) and approximated (dotted lines) α–stable distribution ($G = 4$).*

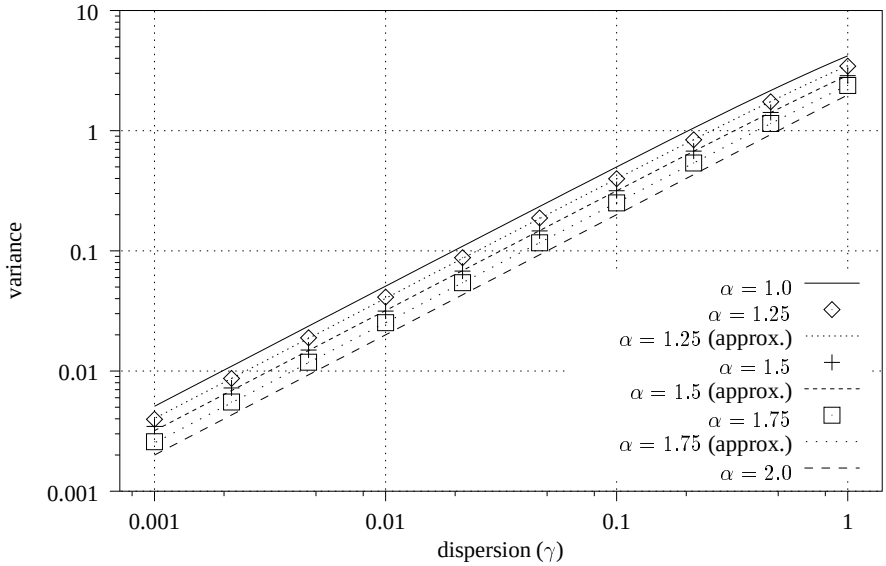


Figure 6.12: *True and approximated variance with respect to the noise dispersion γ ($G = 4$).*

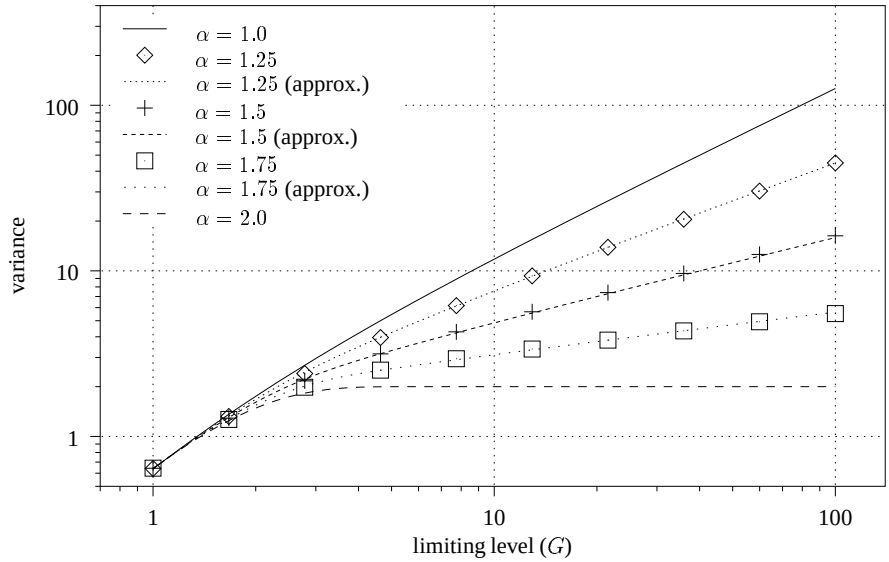


Figure 6.13: *True and approximated variance with respect to the limiting level G ($\gamma = 1$).*

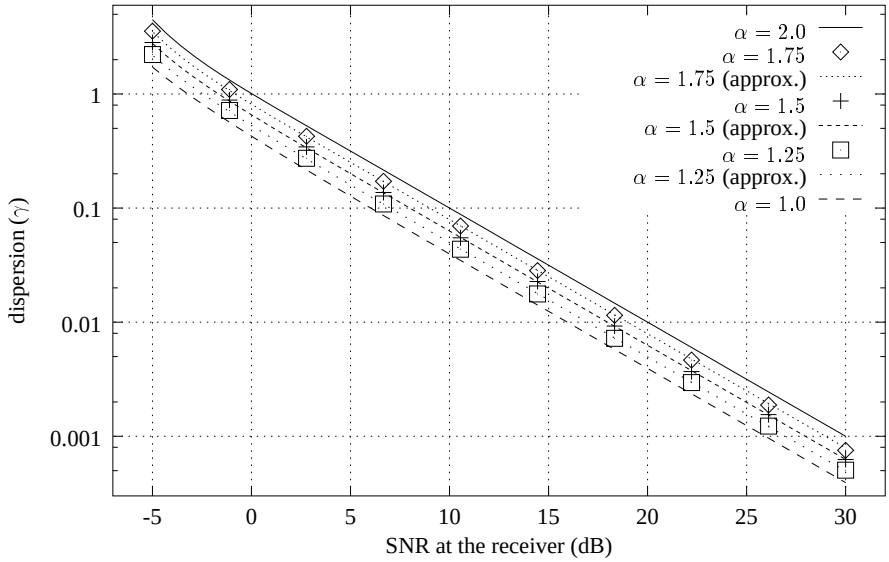


Figure 6.14: *True and approximated noise dispersion γ ($G = 4$).*

reliable for a wide range of SNR's.

6.5 Conclusions

This chapter focused on the performance analysis of the adaptive Bayesian equaliser in α–stable noise environments through experimental simulations. The considered adaptive equaliser comprised a) a non–Gaussian Bayesian detector, b) a robust channel estimation algorithm for stable signals, and c) an estimation algorithm for the parameters of the α–stable noise.

The optimum adaptive Bayesian DFE for α–stable noise environments was presented and its performance in a variety of channel scenarios, stationary and Rayleigh fading, was investigated. For the adaptation of the equaliser centers, two generalised channel estimation algorithms were used, namely MLMP and RLP. The novel framework for the quantitative assessment of systems in such infinite power noise environments, which was presented in chapter 4 was used. According to the experimental results, the proposed adaptive equaliser exhibits a significant performance advantage compared with a conventional adaptive Bayesian equaliser, designed under the Gaussian assumption.

Unfortunately, this performance benefit comes at the expense of unaffordable complexity, because the α–stable density can not, in general, be expressed in closed form. Specifically, the equation that defines the MAP equaliser (eq. (4.30), page 50) involves the evaluation of the stable pdf at $M \cdot N_{\mathrm{c}}$ points for each received sample. On the other hand, the method for the calculation of the working SNR in infinite variance environments (section 4.5, page 64) can only be used for $\alpha = 1$ and $\alpha = 2$. Finally, it is desirable in a simulation environment to parametrically set the noise variance γ to a value which corresponds to a given signal to noise ratio. This implies sovling eq. (4.71) for γ, which can only be done for the two special cases $\alpha = 1$ and $\alpha = 2$. The approximations derived in this chapter enable the evaluation of these quantities with little loss in accuracy, offering near–optimum performance with negligible computational surcharge.

Chapter 7
Summary and conclusions

The research carried out for this thesis addressed the problem of adaptive channel equalisation in environments where the corrupting noise exhibits impulsive characteristics. The optimum Bayesian equaliser for such environments was presented and analysed, while appropriate techniques for the estimation of its parameters were proposed and evaluated. In this chapter the main conclusions of the work are presented, and some suggestions for further research are given.

7.1 Summary and achievements of the work

The motivation for the work undertaken was first put forward. Subsequently the basics of a generic digital communication system were described and a convenient discrete–time equivalent model was adopted, while the need for adaptive equalisation of the communication channel was highlighted.

Although a number of impulsive noise models exist in the literature, the class of stable distributions was chosen in this thesis as an appropriate modeling tool for random signals exhibiting impulsive behaviour. The stable law is supported by the generalised central limit theorem. The family of stable distributions exhibit tails that are heavier than those of the Gaussian density, and the shape of these tails is controlled by a single parameter. No closed–form expressions exist for the stable distributions, while all non–Gaussian stable processes have infinite variance. These features have pivotal consequences in the study of signals and systems using traditional techniques, as most results require the noise pdf in closed form and many techniques are based on second order moments and quadratic criteria.

For the study of systems experiencing infinite variance noise, a novel approach was proposed, which examines the signals after being passed through a finite dynamic range system. This scheme, while maintaining the flexibility and theoretical completeness of the stable law, at

the same time overcomes the problems emerging from the infinite variance of the underlying signals and keeps the validity of many traditional signal processing methods.

However, adopting a non–Gaussian noise model inherently invalidates the optimality of signal processing algorithms designed under the Gaussian assumption. In particular, the Bayesian equaliser designed for Gaussian noise environments is not optimum when the noise exhibits non–Gaussian characteristics. Thus, the maximum a posteriori probabilities criterion was revisited and the optimum symbol–by–symbol equaliser for α–stable noise environments was derived. A very desirable feature of the proposed equaliser is that it is a direct generalisation of the traditional Bayesian equaliser. Moreover, an implementation scheme for this equaliser was given, which turns out to be a direct generalisation of the RBF based implementation of the traditional Bayesian equaliser.

Furthermore, a novel evaluation framework for systems operating in infinite variance environments was presented, based on the finite dynamic range receiver. Under this scheme, the family of α–stable distributions is only used to describe the received signal within the dynamic range of the receiver, thus enabling the modeling of an even broader family of impulsive distributions. Only limited analytical tools were given at this stage, because it is impossible to derive general closed–form expressions involving stable distributions. In particular, the variance of an α–stable signal passed through an ideal saturation device was calculated for $\alpha = 1$ and $\alpha = 2$. However, certain practical approximations were presented later in the thesis to enable the efficient utilisation of the proposed method for any value of α in the interval $[1,\ 2]$, with little loss in accuracy.

The experimental results concerning the Bayesian equaliser in an α–stable noise environment showed a substantial performance deviation between the proposed equaliser and a traditional equaliser in environments where the noise statistics deviate from the Gaussian distribution. However, for these experiments the equalisers had perfect knowledge of the channel and noise characteristics, so the effects from inaccurate estimation of the equaliser parameters were not revealed.

The estimation of the equaliser parameters was subsequently studied; in particular, the problem of estimating the channel noise–free vector states $\mathbf{c}_i$ with their associated signs s_i, and the shape parameters of the α–stable distribution (α and γ) was addressed.

There are two main techniques for the estimation of the equaliser centers. In the indirect method

the centers are inferred from an estimate of the channel impulse response. The direct method, on the other hand, applies a clustering algorithm on the received data. Both schemes were studied, and the means for providing robust behaviour in non–Gaussian noise environments were investigated.

For the indirect method, a unified framework for a variety of robust linear estimation algorithms for non–Gaussian signals was presented. The staring point for this framework was the formulation of channel estimation as an optimisation problem with general optimisation criteria. Although only the quadratic criterion leads to an analytically tractable solution, it is possible to build a recursive process, very similar to the familiar recursive least squares scheme, which can approximate the optimum solution to any general optimisation criterion.

On the other hand, robust versions of traditional clustering algorithms were proposed in the direction of the direct estimation of the equaliser centers. Furthermore, a least p–norm approach was taken as an alternative to the arithmetic averaging (clustering) algorithms. In particular, a robust location estimation algorithm was derived.

The problem of estimating the noise parameters was also discussed, and a comparative presentation of the existing estimation algorithms was conducted. Based on arguments ranging from complexity and efficiency to performance and simplicity, the best candidate for the equalisation problem was found to be the $\log$ FLOM algorithm proposed by Ma and Nikias [124].

Finally, the complete adaptive Bayesian equaliser was experimentally evaluated in an α–stable noise environment. The considered adaptive equaliser consisted of a) a non–Gaussian Bayesian detector, b) a robust channel estimation algorithm for stable signals, and c) an estimation algorithm for the parameters of the α–stable noise.

The experimental results demonstrated that a significant proportion of the promised performance benefit from the optimum adaptive Bayesian equaliser (compared with a conventional adaptive Bayesian equaliser) can still be attained even when the parameters of the equaliser are estimated directly from the received data.

The fact that there is no general closed form expression for the family of α–stable density functions results in unaffordable complexity for the proposed equaliser. However, it was demonstrated that utilising certain approximations it is possible to achieve near–optimum performance with negligible computational overhead, compared to the traditional adaptive Bayesian equal-

iser.

7.2 Limitations of the work and scope for further research

The contributions made by this study have shown that the adaptive Bayesian equaliser designed optimally for α–stable noise environments can provide a substantial performance benefit compared with a Bayesian equaliser designed under the Gaussian assumption. However, certain simplifications and assumptions have been adopted throughout this study, which limit the conclusions of this work from providing a complete understanding of non–Gaussian Bayesian equalisers.

The study of the Bayesian equaliser in this thesis was restricted to the binary transmission scheme. However, this is not always the case in some communication systems, where multi–level transmission constellations are used in order to increase the bit–rate. Although the results of this thesis are indicative, future research on the generalisation of the optimum Bayesian equaliser in non–Gaussian stable noise environments for multi–level transmission schemes would be of great importance.

The experimental results presented in this thesis have been obtained in environments where the channel noise was artificially generated according to the family of α–stable distributions. In that manner, the noise model adopted during the design of the equaliser was accurate. The discussion in section 6.4 indicates that the expected performance of the proposed equaliser may not be substantially inferior, compared with an equaliser designed optimally for any specific impulsive noise model. However, it would be interesting, for future research, to evaluate the performance of the proposed equaliser in a variety of noise environments, generated with diverse impulsive noise models.

Furthermore, the analysis carried out for the incorporation of the infinite variance stable law in a signal processing framework was based on an ideal saturation device at the front end of the receiver. Although this approach was taken in order to simplify the algebra involved, it certainly leads to inaccuracies when a more realistic model (such as the error function type) is adopted for the limiter. Further research can show if these inaccuracies are significant and in what extent they affect the overall performance of the equaliser.

On the other hand, the experiments for the evaluation of the proposed adaptive Bayesian equal-

iser have been performed for linear channel scenarios only. However the scope for the assessment of this equaliser in the presence of non–linear impairments in the channel is still open for research. According to section 4.2, the actual spatial distribution and alignment of the centers largely affects the discrepancies between the optimum α–stable decision boundary and the traditional Gaussian boundary. Accordingly, future work in this direction may demonstrate that the performance benefit from the proposed equaliser is even higher in such environments where the spatial distribution of the centers is less likely to exhibit alignment along the axes, due to the presence of non–linearities.

Finally, the complete adaptive equaliser was studied with the indirect method for the estimation of the centers only, since this technique is more appropriate for linear channels, requires less training data and exhibits better tracking performance [101]. Nevertheless, it would be interesting to evaluate the use of the proposed generalisations of the direct clustering scheme for the estimation of the equaliser centers both in linear and non–linear channel constellations.

References

[1] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*. Cambridge, USA: Cambridge University Press, 2nd ed., 1992.

[2] H. Nyquist, "Certain topics in telegraph transmission theory," *Transactions of AIEE*, vol. 47, pp. 617–644, April 1928.

[3] C. E. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, pp. 379–423, July 1948.

[4] C. E. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, pp. 623–656, October 1948.

[5] K. Abend and B. Fritchman, "Statistical detection for communication channels with intersymbol interference," *Proceedings of the IEEE*, vol. 58, pp. 779–785, 1970.

[6] B. Mulgrew, "Applying radial basis functions," *IEEE Sig. Proc. Magazine*, pp. 50–65, March 1996.

[7] X. Wang and H. Poor, "Robust multi–user detection in non–Gaussian channels," *IEEE Transactions on Signal Processing*, vol. 47, pp. 289–305, February 1999.

[8] D. Middleton, "Non–Gaussian noise models in signal processing for telecommunications: New methods and results for class A and class B noise models," *IEEE Transactions on Information Theory*, vol. 45, pp. 1129–1149, May 1999.

[9] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proceedings of the IEEE*, vol. 81, pp. 986–1009, July 1993.

[10] J. Seo, S. Cho, and K. Feher, "Impact of non–Gaussian impulsive noise on the performance of high–level QAM," *IEEE Transactions on Electromagn. Compat.*, vol. EMC–31, pp. 177–180, May 1989.

[11] L. Izzo and L. Paura, "Error probability for fading CPSK signals in Gaussian and impulsive atmospheric noise environments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES–17, pp. 719–722, September 1981.

[12] W. Henkel and T. Kessler, "Statistical description and modelling of impulsive noise on the German telephone network," *Electronics Letters*, vol. 30, pp. 935–936, June 1994.

[13] K. Blackard, T. Rappaport, and C. Bostian, "Measurements and models of radio frequency impulsive noise for indoor wireless communications," *IEEE J. Select. Areas Commun.*, vol. 11, pp. 991–1001, September 1993.

[14] T. Blankenship, D. Krizman, and T. Rappaport, "Measurements and simulation of radio frequency impulsive noise in hospitals and clinics," *Proc. 1997 IEEE Veh. Technol. Conf. (VTC'97)*, pp. 1942–1946, 1997.

[15] E. Wegman, S. Schwartz, and J. Thomas, *Topics in non–Gaussian Signal Processing*. A Dowen & Culver Book: Springer–Verlag, 1989.

[16] J. G. Proakis, *Digital Communications*. McGraw–Hill, 3rd ed., 1995.

[17] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo–Codes," *IEEE Transactions on Communications*, vol. 44, pp. 1261–1271, October 1996.

[18] R. Steel(Ed), *Mobile Radio Communication*. Pentec Press, London, 1992.

[19] S. U. H. Qureshi, "Adaptive equalization," *Proceedings of the IEEE*, vol. 73, pp. 1349–1387, September 1985.

[20] J. G. D. Forney, "Maximum–likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Transactions on Information Theory*, vol. 18, pp. 363–378, May 1972.

[21] J. G.D. Forney, "The Viterbi algorithm," *Proceedings of the IEEE*, vol. 61, pp. 268–278, March 1973.

[22] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. John Wiley and Sons, 1973.

[23] S. Chen, B. Mulgrew, and S. McLaughlin, "Adaptive Bayesian equalizer with decision feedback," *IEEE Transactions on Signal Processing*, vol. 41, pp. 2918–2927, September 1993.

[24] K. Giridhar, J. J. Shynk, A. Mathur, S. Chari, and R. P. Gooch, "Nonlinear techniques for the joint estimation of cochannel signals," *IEEE Transactions on Communications*, vol. 45, pp. 473–484, April 1997.

[25] R. A. Iltis, J. J. Shynk, and K. Giridhar, "Bayesian algorithms for blind equalisation," *IEEE Transactions on Communications*, vol. 42, pp. 1019–1032, Feb/Mar/Apr 1994.

[26] J. Cid-Sueiro, A. Artés-Rodríguez, and A. Figueiras-Vidal, "Recurrent radial basis function networks for optimal symbol–by–symbol equalisation," *Signal Processing*, vol. 40, pp. 53–63, 1994.

[27] R. Arnott, "Diversity combining for digital mobile radio using radial basis function networks," *Signal Processing*, vol. 63, pp. 1–16, November 1997.

[28] B. Mulgrew, "Nonlinear signal processing for adaptive equalisation and multi–user detection," in *Proceedings of the European Signal Processing Conference*, EUSIPCO, (Island of Rhodes, Greece), pp. 537–544, 8-11 September 1998.

[29] J. Hayes, T. Cover, and J. Riera, "Optimal sequence detection and optimal symbol–by–symbol detection: Similar algorithms," *IEEE Transactions on Communications*, vol. COM–30, pp. 152–157, January 1982.

[30] R. Tanner, D. G. M. Cruickshank, C. Z. W. Hassel–Sweatman, and B. Mulgrew, "Receivers for nonlinearly separable scenarios in DS-CDMA," *Electronics Letters*, vol. 33, pp. 2103–2105, December 1997.

[31] S. Theodoridis, C. Cowan, C. Callender, and C. See, "Schemes for equalisation of communication channels with nonlinear impairments," *IEE Proc. Commun.*, vol. 142, pp. 165–171, June 1995.

[32] J. Cid-Sueiro and A. R. Figueiras-Vidal, "Channel equalization with neural networks," in *Digital Signal Processing in Telecommuniocations - European Project COST#229 Technical Contributions* (A. R. Figueiras-Vidal, ed.), pp. 257–312, London, U.K.: Springer–Verlag, 1996.

[33] S. Patra, B. Mulgrew, and P. Grant, "Subset centre selection with fuzzy implementated radial basis function equaliser design," *CSDSP*, 1998.

[34] B. Mulgrew, P. M. Grant, and J. S. Thompson, *Digital Signal Processing: Concepts and Applications*. Houndmills, Basingstoke, U.K.: Macmillan, 1st ed., 1999.

[35] S. Haykin, *Adaptive Filter Theory*. Prentice–Hall, 3rd ed., 1997.

[36] G. A. Tsihrintzis and C. L. Nikias, "Performance of optimum and suboptimum receivers in the presence of impulsive noise modelled as an alpha–stable process," *Proceedings MILCOM '93*, pp. 658–662, October 1993.

[37] L. Lind and N. Mufti, "Efficient method for modelling impulse noise in a communication system," *Electronics Letters*, vol. 32, pp. 1440–1441, August 1996.

[38] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw–Hill, 1st ed., 1965.

[39] D. Middleton, "Man–made noise in urban environments and transportation systems: Models and measurement," *IEEE Transactions on Communications*, vol. COMM–21, pp. 1232–1241, 1973.

[40] D. Middleton, "Statistical–physical models of electromagnetic interference," *IEEE Transactions on Electromagn. Compat.*, vol. EMC–19, pp. 106–127, August 1977.

[41] P. Brockett, M. Hinich, and G. Wilson, "Nonlinear and non–Gaussian ocean noise," *J. Acoust. Soc. Amer.*, vol. 82, pp. 1286–1399, 1987.

[42] D. Middleton, "Channel modelling and threshold signal processing in under–water acoustics: An analytical overview," *IEEE J. Oceanic Eng.*, vol. OE–12, pp. 4–28, 1987.

[43] P. Lévy, *Probabilités*. Paris: Gauthier–Villars, 1925.

[44] A. Weron, "Stable processes and measures: A survey," in *Probability Theory on Vector Spaces III* (S. Cambanis, ed.), pp. 306–364, Berlin: Springer, 1983.

[45] A. Weron and K. Weron, "Stable measures and processes in statistical physics," in *Probability Theory on Vector Spaces IV* (S. Cambanis, ed.), pp. 440–452, Berlin: Springer, 1987.

[46] V. Zolotarev, *One-Dimensional Stable Distributions*. American Mathematical Society, 1986.

[47] J. Holtsmark, "Über die Verbreiterung von Sektrallinien," *Ann. Physik*, vol. 58, no. 4, pp. 577–630, 1919.

[48] B. Mandelbrot, "The variation of certain speculative prices," *J. of Business*, vol. 36, pp. 394–419, 1963.

[49] E. Fama, "The behavior of stock–market prices," *J. of Business*, vol. 38, pp. 34–105, 1965.

[50] S. Press, "A compound events model for security prices," *J. of Business*, vol. 40, pp. 317–335, 1968.

[51] J. Teichmöller, "A note on the distribution of stock price changes," *Journal of the American Statistical Association*, vol. 66, pp. 282–284, 1971.

[52] B. Mandelbrot and J. V. Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, pp. 422–437, 1968.

[53] J. Berger and B. Mandelbrot, "A new model for error clustering in telephone circuits," *IBM J. Research and Development*, vol. 7, pp. 224–236, 1963.

[54] M. Shao and C. Nikias, "A symmetrical stable model for impulsive noise," technical report USC-SIPI-231, University of Southern California, February 1993.

[55] S. Rappaport and L. Kurz, "An optimal nonlinear detector for digital data transmission throug non–Gaussian channels," *IEEE Trans. Comm. Techn.*, vol. COM–14, pp. 266–274, 1966.

[56] B. Stuck and B. Kleiner, "A statistical analysis of telephone noise," *Bell. Syst. Tech. J.*, vol. 53, no. 7, pp. 1263–1320, 1974.

[57] J. M. Chambers, C. L. Mallows, and B. W. Stuck, "A method for simulating stable random variables," *Journal of the American Statistical Association*, vol. 71, pp. 340–344, June 1976.

[58] S. Bates, *Traffic characterisation and modelling for call admission control schemes on asynchronous transfer mode networks*. PhD thesis, Department of Electronics & Electrical Eng., Edinburgh University, 1997.

[59] G. Samorodnitsky and M. Taqqu, *Stable non–Gaussian Random Processes*. Chapman & Hall, 1994.

[60] W. Feller, *An Introduction to Probability Theory and its Applications*. New York: Wiley, 1971.

[61] E. Lukacs, *Characteristic Functions*. London: Griffin, 1960.

[62] D. Holt and E. Crow, "Tables and graphs of the stable probability density functions," *J. of Research Nat. Bureau of Standard - B. Mathematical Sciences*, vol. 77B, no. 3 & 4, pp. 143–198, 1973.

[63] H. Bergstrom, "On some expansions of stable distribution functions," *Ark. Math.*, vol. 2, pp. 375–378, 1952.

[64] L. Breiman, *Probability*. MA: Addison–Wesley, 1968.

[65] S. Cambanis, *Probability Theory on Vector Spaces III*. Berlin: Springer, 1983.

[66] C. Granger and D. Orr, ""infinite variance" and research strategy in time series analysis," *J. Amer. Statist. Soc.*, vol. 67, pp. 275–285, June 1972.

[67] P. Bloomfield, *Least Absolute Deviations: Theory*. Boston: Birkhäuser, 1983.

[68] R. Gonin and A. Money, *Nonlinear L_p–Norm Estimation*. New York: Marcel Dekker, 1989.

[69] A. Tarantola, *Inverse Problem Theory*. New York: Elsevier, 1987.

[70] D. Cline and P. Brockwell, "Linear prediction of ARMA processes with infinite variance," *Stochastic Processes and Theory*, vol. 19, pp. 281–296, 1985.

[71] J. Schröder, R. Yarlagadda, and J. Hershey, "Linear predictive spectral estimation via the l_1 norm," *Signal Processing*, vol. 17, pp. 19–29, 1989.

[72] R. Yarlagadda, J. Bednar, and T. Watt, "Fast algorithms for l_p deconvolution," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, pp. 174–182, 1985.

[73] J. Schröder, R. Yarlagadda, and J. Hershey, "l_p normed minimisation with applications to linear predictive modelling for sinusoidal frequency estimation," *Signal Processing*, vol. 24, pp. 193–216, 1991.

[74] S. Cambanis and G. Miller, "Linear problems in pth order and stable processes," *SIAM J. Appl. Math*, vol. 41, pp. 43–69, August 1981.

[75] M. Schilder, "Some structure theorems for the symmetric stable laws," *Ann. Math. Statist.*, vol. 41, pp. 412–421, 1970.

[76] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalisation using radial basis function networks," *IEEE Transactions on Neural Networks*, vol. 4, pp. 570–579, July 1993.

[77] A. T. Georgiadis and B. Mulgrew, "A MAP equaliser for impulsive noise environments," *First IMA International Conference on Mathematics in Communications*, pp. 81–86, December 1998.

[78] S. Siu, G. Gibson, and C. Cowan, "Decision feedback equalisation using neural network structures and performance comparison with the standard architecture," *IEE Proceedings* Part I, vol. 137, pp. 221–225, August 1990.

[79] T. Poggio and F. Girosi, "Network for approximation and learning," *Proceedings of the IEEE*, vol. 78, pp. 1481–1497, September 1990.

[80] S. Chen, B. Mulgrew, E. Chng, and G. Gibson, "Space translation properties and the minimum–BER linear–combiner DFE," *IEE Proc.–Commun.*, vol. 145, pp. 316–322, October 1998.

[81] N. Dyn, "Interpolation of scattered data by radial functions," in *Topics in Multivariate Interpolation* (C. K. Chiu, L. L. Schumaker, and F. Ultreras, eds.), pp. 47–61, New York: Academic Press, 1987.

[82] M. J. D. Powell, "Radial basis function for multivariable interpolation: A review," in *Algorithms for Approximation of Functions and Data* (J. C. Mason and M. G. Cox, eds.), pp. 143–167, Oxford University Press, 1987.

[83] C. A. Micchelli, "Interpolation of scattered data: Distance matrix and conditionally positive functions," *Constructive Approximation*, vol. 2, no. 1, pp. 11–22, 1986.

[84] D. S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Systems*, vol. 2, pp. 321–355, 1988.

[85] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithms for radial basis function networks," *IEEE Transactions on Neural Networks*, vol. 2, pp. 302–309, March 1991.

[86] E.-S. Chng, H. Yang, and W. Skarbek, "Reduced complexity implementation of bayesian equalizer using local RBF network for channel equalization problem," *Electronics Letters*, vol. 32, pp. 17–19, January 1996.

[87] S. Chen, S. A. Billings, C. F. N. Cowan, and P. M. Grant, "Non–linear system identification using radial basis functions," *International Journal of Systems Science*, vol. 21, pp. 2513–2539, December 1990.

[88] S. Chen, S. A. Billings, and P. M. Grant, "Recursive hybrid algorithm for non–linear system identification using radial basis function networks," *International Journal of Control*, vol. 55, pp. 1051–1070, May 1992.

[89] I. Cha and S. A. Kassam, "Interference cancellation using radial basis function network," *Signal Processing (Eurasip)*, vol. 47, pp. 247–268, December 1995.

[90] Q. Zhao and Z. Bao, "Radar target recognition using a radial basis function neural network," *Neural Networks, Elsevier Science*, vol. 9, no. 4, pp. 709–720, 1996.

[91] J. A. Leonard and M. A. Kramer, "Radial basis function networks for classifying process faults," *IEEE Communications Systems Magazine*, pp. 31–38, April 1991.

[92] L. Tarassenko and S. Roberts, "Supervised and unsupervised learning in radial basis function classifiers," *IEE Proceedings - Vision Image Signal Processing*, vol. 141, pp. 210–216, August 1994.

[93] S. Chen, G. J. Gibson, C. F. N. Cowan, and P. M. Grant, "Reconstruction of binary signals using an adaptive radial basis function equalizer," *Signal Processing (Eurasip)*, vol. 22, pp. 77–93, January 1991.

[94] S. K. Patra and B. Mulgrew, "Computational Aspects for Adaptive Radial Basis Function Equalizer Design," in *Proceedings of IEEE International Symposium on Circuits and Systems*, vol. 1, (Hong Kong), pp. 521–524, IEEE, 9-12 June 1997.

[95] M. A. Tugay and Y. Tanik, "Properties of the momentum LMS algorithm," *Signal Processing (Eurasip)*, vol. 18, no. 2, pp. 117–127, 1989.

[96] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares algorithms for training multioutput radial basis function networks," *Proceedings-F of the IEE : Radar and Signal Processing*, vol. 139, pp. 378–384, December 1992.

[97] S. Chen, S. McLaughlin, and B. Mulgrew, "Complex–valued radial basis function network, part I: Network architecture and learning algorithms," *Signal Processing (Eurasip)*, vol. 35, pp. 19–31, January 1994.

[98] S. Chen, S. McLaughlin, and B. Mulgrew, "Complex-valued radial basis function network, part II: Application to digital communication channel equalization," *Signal Processing (Eurasip)*, vol. 36, pp. 175–188, March 1994.

[99] A. Clark, L. Lee, and R. Marshall, "Developments of the conventional nonlinear equaliser," *IEE Proccedings*, vol. 129, no. 2, pp. 85–94, 1982.

[100] D. Williamson, R. A. Kennedy, and G. W. Pulford, "Block decision feedback equalization," *IEEE Transactions on Communications*, vol. 40, pp. 255–264, February 1992.

[101] S. Chen, S. McLaughlin, B. Mulgrew, and P. Grant, "Adaptive bayesian decision feedback equaliser for dispersive mobile radio channels," *IEEE Transactions on Communications*, vol. 43, pp. 1937–1945, May 1995.

[102] B. Mulgrew and C. Cowan, *Adaptive Filters and Equalisers*. Kluwer Academic Publishers, 1988.

[103] O. Arikan, A. Çetin, and E. Erzin, "Adaptive filtering for non–Gaussian stable processes," *IEEE Signal Processing Letters*, vol. 1, pp. 163–165, November 1994.

[104] H. Dai and N. Sinha, "Robust recursive least–squares method with modified weights for bilinear system identification," *Proceedings of the IEE*, vol. 136, pp. 122–126, May 1989.

[105] E. Kuruoğlu, W. Fitzgerald, and P. Rayner, "Non–linear autoregressive modeling of non–Gaussian signals using L_p–norm techniques," *International Conference on Acoustics, Speech and Signal Processing*, vol. 3, pp. 3533–3536, 1997.

[106] P. Huber, *Robust Statistics*. New York: Wiley, 1981.

[107] A. T. Georgiadis and B. Mulgrew, "A family of recursive algorithms for channel identification in alpha–stable noise," *5th Bayona workshop on emerging technologies in telecommunications*, pp. 153–157, September 1999.

[108] J. Proakis, *Advanced Digital Signal Processing*. New York: Macmillan, 1992.

[109] J. Bodenschatz and C. Nikias, "Recursive local orthogonality filtering," *IEEE Transactions on Signal Processing*, vol. 45, pp. 2293–2300, September 1997.

[110] S. Puthenpura, N. Sinha, and O. Vidal, "Application of M–estimation in robust recursive system identification," *IFAC Symp. Stocastic Control*, pp. 23–30, 1985.

[111] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. MIT Press, 1983.

[112] R. H. Byrd and D. A. Payne, "Convergence of the iteratively reweighted least squares algorithm for robust regression," technical report 313, The Johns Hopkins Univ. Baltimore, MD, June 1979.

[113] I. Pitas and A. Venetsanopoulos, *Nonlinear Digital Filters : Principles and Applications*. Kluwer International Series in Engineering and Computer Science. Vlsi, Computer Architecture, Kluwer Academic Publishers, 1990.

[114] W. DuMouchel, *Stable Distributions in Statistical Inference*. PhD thesis, Department of Statistics, Yale University, 1971.

[115] B. Brorsen and S. Yang, "Maximum likelihood estimates of symmetric stable distribution parameters," *Commun. Statist.–Simula.*, vol. 19, no. 4, pp. 1459–1464, 1990.

[116] J. McCulloch, "Simple consistent estimators of stable distribution parameters," *Commun. Statist. – Simula.*, vol. 15, no. 4, pp. 1109–1136, 1986.

[117] E. Fama and R. Roll, "Parameter estimates for symmetric stable distributions," *Journal of the American Statistical Association*, vol. 66, pp. 331–338, June 1971.

[118] E. Fama and R. Roll, "Some properties of symmetric stable distributions," *Journal of the American Statistical Association*, vol. 63, pp. 817–836, September 1968.

[119] S. Press, "Estimation in univariate and multivariate stable distributions," *Journal of the American Statistical Association*, vol. 67, pp. 842–846, December 1972.

[120] A. Paulson, E. Holcomb, and R. Leitch, "The estimation of the parameters of the stable laws," *Biometrika*, vol. 62, pp. 163–170, 1975.

[121] I. Koutrouvelis, "Regression–type estimation of the parameters of stable laws," *Journal of the American Statistical Association*, vol. 75, pp. 918–928, December 1980.

[122] I. Koutrouvelis, "An iterative procedure for the estimation of the parameters of stable laws," *Commun. Statist. – Simula.*, vol. 10, no. 1, pp. 17–28, 1981.

[123] V. Akgiray and C. Lamoureux, "Estimation of stable–law parameters: A comparateive study," *J. Business and Economic Statistics*, vol. 7, pp. 85–93, January 1989.

[124] X. Ma and C. L. Nikias, "Parameter estimation and blind channel identification in impulsive signal environments," *IEEE Transactions on Signal Processing*, vol. 43, pp. 2884–2897, December 1995.

[125] G. A. Tsihrintzis and C. L. Nikias, "Fast estimation of the parameters of alpha–stable impulsive interference using asymptotic exterme value theory," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pp. 1840–1843, 1995.

[126] G. A. Tsihrintzis and C. L. Nikias, "Fast estimation of the parameters of alpha–stable impulsive interference," *IEEE Transactions on Signal Processing*, vol. 44, pp. 1492–1503, jun 1996.

[127] S. Chen and B. Mulgrew, "Minimum–SER linear–combiner decision feedback equaliser," *IEE Proc.–Commun.*, vol. 146, pp. 347–353, December 1999.

[128] W. C. Jakes, ed., *Microwave Mobile Communications*. New York: IEEE Press, 1993.

[129] S. O. Rice, "Mathematical analysis of random noise," *Bell Systems Technical Journal*, vol. 23, pp. 282–332, July 1944.

[130] S. O. Rice, "Mathematical analysis of random noise," *Bell Systems Technical Journal*, vol. 24, pp. 46–156, January 1945.

[131] M. Pätzold, U. Killat, F. Laue, and Y. Li, "On the statistical properties of deterministic simulation models for mobile fading channels," *IEEE Transactions on Vehicular Technology*, vol. 47, pp. 254–269, February 1998.

Appendix A
Publications

The following papers have been either published or submitted to conferences or journals.

A. T. Georgiadis and B. Mulgrew "A MAP equaliser for impulsive noise environments," *First IMA International Conference on Mathematics in Communications*, December 1998.

A. T. Georgiadis and B. Mulgrew "A family of recursive algorithms for channel identification in alpha–stable noise," *5th Bayona workshop on emerging technologies in telecommunications*, pp. 153–157, September 1999.

A. T. Georgiadis and B. Mulgrew "Adaptive Bayesian decision feedback equaliser for alpha-stable noise environments," Submitted to *: Signal Processing – Special Issue on "Signal Processing Techniques for Emerging Communications Applications"*, 2000.

A MAP Equaliser for Impulsive Noise Environments

A.T. Georgiadis and B. Mulgrew
Department of Electronic & Electrical Engineering
University of Edinburgh, Scotland, UK

Abstract

The maximum a-posteriori probabilities (MAP) criterion for channel equalisation has been used with the assumption of additive Gaussian white noise. In this study, we investigate the MAP criterion in the presence of finite inter-symbol interference and additive impulsive noise modelled as an α-stable process. The optimum symbol-by-symbol Bayesian detector is presented and its performance is then discussed. The experimental results suggest that the proposed estimator outperforms the traditional Bayesian estimator based on the Gaussian noise assumption.

1. Introduction

The Gaussian process has been always the dominant noise model in communications and signal processing literature, mainly because of the central limit theorem. In addition, the Gaussian assumption often leads to analytically tractable solutions [1]. Unfortunately, in many communication channels, the observation noise exhibits Gaussian, as well as impulsive[1] characteristics. The sources of impulsive noise may be either natural (e.g. lightnings), or man made. It may include atmospheric noise or ambient acoustic noise. It might come from relay contacts, electro-magnetic devices, electronic apparatus, or transportation systems, switching transients, and accidental hits in telephone lines [2, 3].

Most of the systems are optimised under the Gaussian assumption and their performance is significantly degraded by the occurrence of impulsive noise [4, 5]. That is, more realistic statistical models must be used [1]. Impulsive noise is more likely to exhibit sharp spikes or occasional bursts of outlying observations than one would expect from normally distributed signals. The empirical data indicates that the probability density functions (pdf's) of the impulsive noise processes exhibit a similarity to the Gaussian pdf, being bell-shaped, smooth, and symmetric, but at the same time having significantly heavier tails [2]. A variety of impulsive noise models have been proposed (e.g. [3, 6]). Recently, however, it has been suggested [1] that the family of α-stable random variables provides useful models for impulsive phenomena.

2. The class of α-stable random variables

The stable law is a direct generalisation of the Gaussian distribution and in fact includes the Gaussian as a limiting case. In general, stable distributions are usually assumed to be non-Gaussian although the Gaussian distribution belongs to the family of stable distributions.

[1] the term "impulsive" is used to indicate the probability of large interference levels

The main difference between non-Gaussian stable distribution and Gaussian distribution is that the tails of the stable density are heavier than those of the Gaussian density. This characteristic of the stable distribution is one of the main reasons why it is suitable for modelling signals and noise of impulsive nature. The α-stable distribution $\phi_{\mathbf{x}}(x)$ is defined by means of its characteristic function [2] [1]

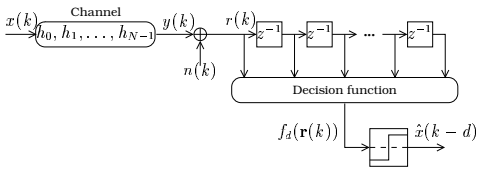


Figure 1. The communication system model

$$\Phi_{\mathbf{x}}(\omega) = \exp\left(j\delta\omega - \gamma|\omega|^{\alpha}[1 + j\beta sgn(\omega)\rho(\omega,\alpha)]\right)$$

where

$$\rho(\omega,\alpha) = \begin{cases} \tan\frac{\alpha\pi}{2} & : \quad \alpha \neq 1 \\ \frac{2}{\pi}\log|\omega| & : \quad \alpha = 1 \end{cases}$$

The α-stable distribution is completely determined by four parameters: 1) the characteristic exponent α, controlling the heaviness of the pdf tails, 2) the index of skewness β, controlling the symmetry of the pdf, 3) the scale parameter γ, also called the *dispersion* [3] , and 4) the location parameter δ, under the restrictions:

$$0 < \alpha \leq 2, \quad -1 \leq \beta \leq 1, \quad \gamma > 0, \quad -\infty < \delta < \infty.$$

Unfortunately, there is no closed expression for $\phi(x)$, except $\alpha = 1$ and $\alpha = 2$ cases, which correspond to the Cauchy and Gaussian distributions, respectively. In practice, however, it is convenient to assume that the α-stable pdf is symmetric ($\beta = 0$) around 0 ($\delta = 0$):

$$\Phi(\omega) = \exp\left(-\gamma|\omega|^{\alpha}\right)$$

in which case the characteristic function is real and even. That is, the pdf can be simplified

$$\phi_{\mathbf{x}}(x) = \frac{1}{\pi}\int_0^{\infty} \Phi_{\mathbf{x}}(\omega)\cos\omega x dx$$

A very important property of non-Gaussian α-stable distributions is that their second order moments (variance) are infinite.

3. The optimum MAP estimator

We assume that the digital data sequence $\{x(k)\}$ is passed through an N-order, noiseless linear dispersive channel $H(z) = \sum_{i=0}^{N-1} h_i z^{-i}$. The observed sequence $\{r(k)\}$ is formed by adding random noise $\{n(k)\}$ to the output of the channel:

$$y(k) = \sum_{i=0}^{N-1} h_i x(k-i), \quad r(k) = y(k) + n(k)$$

[2] The characteristic function $\Phi_{\mathbf{x}}(\omega)$ of a random variable (r.v.) $\mathbf{x}$ is the Fourier transform of its probability density function $\phi_{\mathbf{x}}(x)$

[3] In the Gaussian case ($\alpha = 2$) the dispersion parameter (γ) equals $\sigma^2/2$.

In symbol-by-symbol equalisation, the observation vector $\mathbf{r}(k) = [r(k), \quad r(k-1), \ldots, \quad r(k-M+1)]^T$ is fed into a decision function $f_d(\mathbf{r}(k))$ to produce an estimate $\hat{x}(k-d)$ of the transmitted bit $x(k-d)$, where d is the equaliser's decision delay (Fig. 1).

Having observed the vector $\mathbf{r}(k)$, we decide $+1$ if the probability that it was caused by $x(k-d) = +1$ exceeds the probability that it was caused by $x(k-d) = -1$, and vice versa. Using Bayes' rule [7], the a-posteriori probabilities may be expressed by means of the a-priori probabilities

$$P(x|\mathbf{r}) = \frac{p(\mathbf{r}|x)P(x)}{p(\mathbf{r})} \tag{1}$$

We can assume, without loss of generality, that the transmitted symbols are equiprobable; we also observe that the denominator in eq. (1) is independent of which signal is transmitted. Consequently, the aforementioned decision rule is equivalent to finding the symbol x that maximises the likelihood of $\mathbf{r}$, that is $p(\mathbf{r}|x)$. Therefore, the decision function of the equaliser should be

$$f_d(\mathbf{r}(k)) = p(\mathbf{r}(k)|\{x(k-d) = +1\}) - p(\mathbf{r}(k)|\{x(k-d) = -1\})$$

with the decision based on the sign of $f_d(\mathbf{r}(k))$.

At this point it is appropriate to highlight that, since the channel is assumed to be a FIR filter, and the transmitted sequence is binary, the noiseless channel output $y(k)$ can only take 2^N possible discrete values [4] $\{y_0, \quad y_1, \quad y_{2^N-1}\}$ The noiseless observation vector $\mathbf{y}(k)$, containing M sequential channel output observations, will exhibit discrete nature, as well. In order to explore the discrete character of the noise-free observation vector, it is convenient to describe $\mathbf{y}(k)$ in matrix form [8]

$$\begin{aligned}
\mathbf{y}(k) &= \mathbf{H} \cdot \mathbf{x}'(k) \\
&= \begin{bmatrix} h_0 & \cdots & h_{N-1} & 0 & \cdots & 0 \\ 0 & \cdots & h_{N-2} & h_{N-1} & \cdots & 0 \\ & & & & & \\ 0 & \cdots & h_0 & h_1 & \cdots & h_{N-1} \end{bmatrix} \cdot \begin{bmatrix} x(k) \\ x(k-1) \\ \cdots \\ x(k-N-M+2) \end{bmatrix}
\end{aligned}$$

That is, there are totally 2^{N+M-1} discrete states $\mathbf{c}_i$ for the noise-free observation vector $\mathbf{y}(k)$. For the simple case of a 2 coefficient channel ($H_1(z) = 1 + 0.5z^{-1}$) there are 8 possible 2-dimensional channel output vectors. Fig. 2 depicts each of these 8 possible points plotted as either a star to indicate that the output vector represents an input $x(k) = -1$ or a circle to represent an input $x(k) = +1$. The set of noise free channel output states $\mathbf{c}_i$ can be partitioned into two sets, conditioned on the transmitted symbol of interest:

$$S^+ = [\mathbf{c}|x(k-d) = +1], \quad S^- = [\mathbf{c}|x(k-d) = -1]$$

Consequently, as shown in [8], we can rewrite the decision function

$$\begin{aligned}
f_d(\mathbf{r}(k)) &= \sum_{\mathbf{c}_i \epsilon S^+} p(\mathbf{r}(k)|\mathbf{c}_i)P(\mathbf{c}_i) - \sum_{\mathbf{c}_i \epsilon S^-} p(\mathbf{r}(k)|\mathbf{c}_i)P(\mathbf{c}_i) \\
&= \frac{1}{N_r} \sum_{\mathbf{c}_i \epsilon S} s_i \phi(\mathbf{r}_1(k)|\mathbf{c}_{i,1}) \cdot \cdots \phi(\mathbf{r}_M(k)|\mathbf{c}_{i,M})
\end{aligned} \tag{2}$$

where

[4] All the possible noiseless channel outputs are usually referred as *scalar centers*

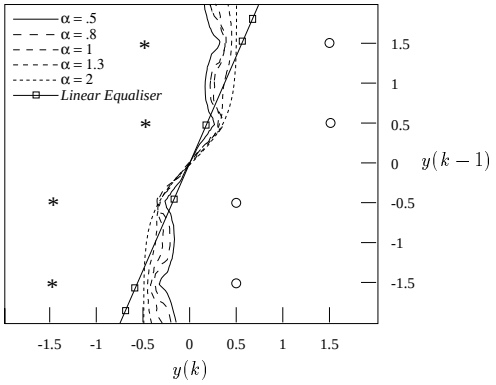


Figure 2. Observation space for channel $H_1(z)$ and the proposed Bayesian boundaries for various values of α. The linear equaliser boundary is also depicted.

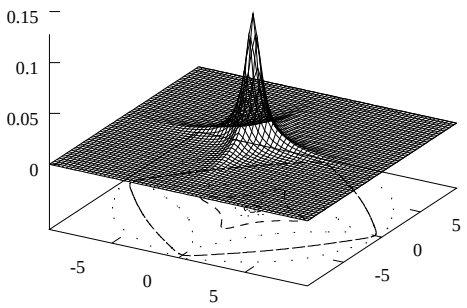


Figure 3. The two dimensional α-stable pdf for $\alpha = 0.8$

$$\mathbf{r}(k) \;=\; [\mathbf{r}_1(k), \quad \mathbf{r}_2(k), \quad \ldots, \quad \mathbf{r}_M(k)]^{\mathrm{T}} \qquad , \qquad s_i = \begin{cases} +1 & : \quad \mathbf{c}_i \epsilon S^+ \\ -1 & : \quad \mathbf{c}_i \epsilon S^- \end{cases}$$

and $\phi(x)$ is the pdf of the additive noise, modelled as an $\alpha-$stable process. Since we are interested merely in the sign of $f_d(\mathbf{r}(k))$, eq. (2) can be simplified

$$f_d(\mathbf{r}(k)) \;=\; \sum_{\mathbf{c}_i \epsilon S} s_i \phi(\mathbf{r}_1(k) - \mathbf{c}_{i,1}) \phi(\mathbf{r}_2(k) - \mathbf{c}_{i,2}) \cdots \phi(\mathbf{r}_M(k) - \mathbf{c}_{i,M}) \qquad (3)$$

The likelihood $p(\mathbf{r}(k)|\mathbf{c}_i)$ exhibits radial symmetry only in $\alpha = 2$ case, resulting from the interesting property of Gaussian pdf's $g(\mathbf{x}) = e^{-\frac{\mathbf{x}_1^2}{2\sigma^2}} \cdot e^{-\frac{\mathbf{x}_2^2}{2\sigma^2}} \cdots e^{-\frac{\mathbf{x}_M^2}{2\sigma^2}} = e^{-\frac{\|\mathbf{x}\|^2}{2\sigma^2}}$, where $\| \mathbf{x} \|$ is the *Euclidean norm* of vector $\mathbf{x}$.

Fig. 3 depicts the $\alpha = 0.8$ case of a two dimensional *SαS* pdf, where the rather *diamond* instead of *radial-symmetric* profile can be clearly seen. This lack of radial symmetry, leads to radical alterations in the decision boundary of a Bayesian equaliser, compared to the Gaussian noise assumption.

4. Experiments

Two second order channel models have been simulated: $H_1(z) = 1 + 0.5z^{-1}$ and $H_2(z) = 0.5 + z^{-1}$. The length of the observation vector was $M = 2$ and the lag of the equaliser is assumed $d = 0$. Fig. 2 depicts the impact of the characteristic exponent α on the optimum detection boundary determined by eq. (2) for channel $H_1(z)$. The decision boundary determined by the MSE linear equaliser (assuming that the noise model is Gaussian) is also depicted.

Fig. 4 shows the observation space for channel $H_2(z)$. The decision boundaries formed by eq. (2) and the traditional Bayesian equaliser, are also shown, respectively. This is a non-minimum phase channel, and it is not linearly separable. Moreover, it is clearly seen that the α-stable noise model assumption leads to radical alterations of the optimum decision boundary compared to that of a traditional Bayesian equaliser. Fig. 5

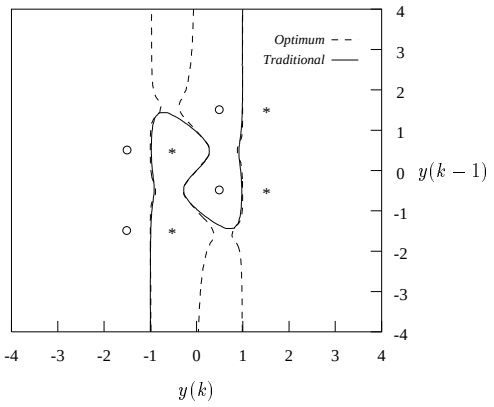


Figure 4. Observation space for channel $H_2(z)$ and the proposed Bayesian decision boundary (α = 1) compared to that of a traditional Bayesian equaliser .

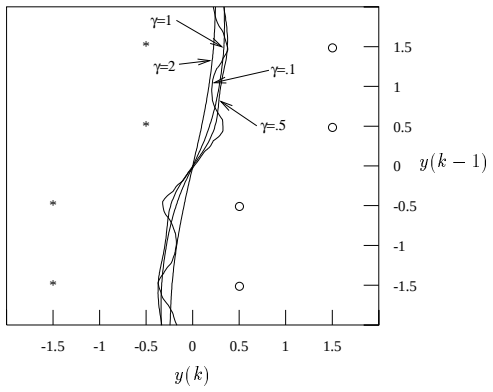


Figure 5. Observation space for channel $H_1(z)$ and the Bayesian decision boundary for various values of γ (α=1).

shows the optimum decision boundaries changing as the dispersion parameter γ increases. For small values of γ the decision boundary is strongly affected by the characteristic exponent α. On the contrary, as γ increases, the decision boundary becomes relatively independent to α and tends to be linear-like.

The traditional performance measures are plots of the bit error-rate (B.E.R.) against the Eb/N_o ratio, where Eb is the normalised bit energy and N_o is the white noise power spectral density ($N_o = \sigma^2/2$ in the Gaussian case). Since the α-stable noise power spectral density is infinite, it is convenient to use the relationship between γ and σ. That is, the generalisation of the ratio Eb/N_o into $Eb/4\gamma$ is intuitively justified, by replacing the noise power spectral density by its dispersion. The performance of the proposed MAP equaliser, based on eq. (2), has been simulated, along with the traditional Bayesian equaliser based on the Gaussian noise assumption and the linear equaliser. The B.E.R. performance against the $Eb/4\gamma$ ratio of all three equalisers is plotted in Fig. 6 (for channel $H_1(z)$ and a variety of values for α). The B.E.R. axis is in logarithmic scale and the ratio $Eb/4\gamma$ is expressed in decibels.

As the characteristic exponent α increases (that is the noise becomes more impulsive) the performance of all three equalisers deteriorate significantly. As the noise becomes stronger (large γ) the decision boundary tends to be linear-like and the performance of the optimum MAP equaliser is close to that of the traditional MAP, as well as the linear equaliser.

The proposed optimum MAP equaliser clearly outperforms the traditional Bayesian estimator, especially when it comes to cope with the non-minimum phase channel $H_2(z)$. The results of the experiments for this channel, for various values of α, are depicted in Fig. 7. The performance of the linear equaliser is severely degraded in this case (non linearly separable channel). The decision boundary determined by the proposed equaliser for this channel is significantly different from that of a traditional Bayesian estimator, as shown in Fig. 4. This can justify the improvement in performance introduced by the proposed equaliser.

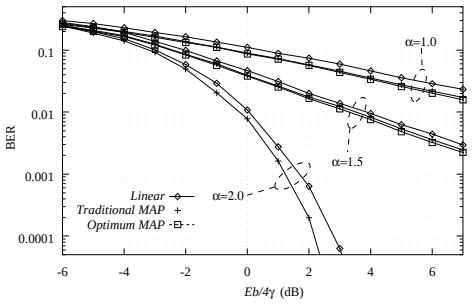


Figure 6. Performance of the Optimum Bayesian, Traditional Bayesian and Linear equaliser for channel $H_1(z)$ for various values of the characteristic exponent α

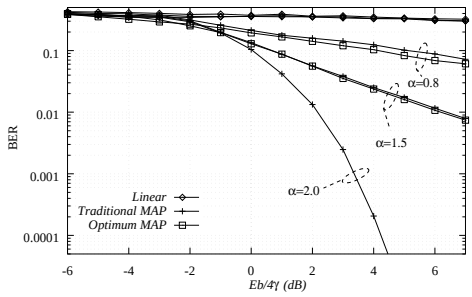


Figure 7. Performance of the Optimum Bayesian, Traditional Bayesian and Linear equaliser for channel $H_2(z)$ for various values of the characteristic exponent α.

5. Conclusions

In this paper, we revised the *maximum a-posteriori probabilities criterion* in the presence of additive impulsive noise modelled as an α-stable process. We then derived the optimum symbol-by-symbol Bayesian (MAP) equaliser in such an impulsive noise environment. The proposed equaliser is implementable. The experimental results suggest that the proposed optimum equaliser outperforms the Bayesian equaliser designed under the Gaussian assumption. The performance gain is due to the changes in the optimum decision boundary which occur as the noise statistics deviate from the Gaussian distribution. Nevertheless, further work is still due to be done on new training algorithms for estimating the centers of the proposed MAP equaliser in impulsive noise environments.

References

[1] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proceedings of the IEEE*, vol. 81, pp. 986–1009, July 1993.

[2] G. A. Tsihrintzis and C. L. Nikias, "Performance of optimum and suboptimum receivers in the presence of impulsive noise modelled as an alpha-stable process," *Proceedings MILCOM '93*, pp. 658–662, October 1993.

[3] L.F.Lind and N.A.Mufti, "Efficient method for modelling impulse noise in a communication system," *IEE Electronics Letters*, May 1996.

[4] J.Seo, S.Cho, and K.Feher, "Impact of non-Gaussian impulsive noise on the performance of high-level QAM," *IEEE Transactions on Electromagn. Compat.*, vol. EMC-31, pp. 177–180, May 1989.

[5] L.Izzo and L.Paura, "Error probability for fading CPSK signals in Gaussian and impulsive atmospheric noise environments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-17, pp. 719–722, September 1981.

[6] W.Henkel and T.Kessler, "Statistical description and modelling of impulsive noise on the German telephone network," *IEE Electronics Letters*, March 1994.

[7] J. G. Proakis, *Digital Communications*, McGraw-Hill, 3 edition, 1995.

[8] B. Mulgrew, "Applying radial basis functions," *IEEE Signal Processing Magazine*, pp. 50–65, March 1996.

A FAMILY OF RECURSIVE ALGORITHMS FOR CHANNEL IDENTIFICATION IN ALPHA-STABLE NOISE

Apostolos Th. Georgiadis and Bernard Mulgrew

Department of Electronics & Electrical Engineering
University of Edinburgh, EH9 3JL Edinburgh, UK
Tel: +44-131-6505655, +44-131-6505580 fax: +44-131-6505580 E-mail: [atg, bernie]@ee.ed.ac.uk

ABSTRACT

In some communication systems the channel noise is known to be non-Gaussian due, largely, to impulsive phenomena. The performance of quadratic signal processing algorithms may degrade seriously in such environments. In this paper we present a family of recursive algorithms for robust adaptive filtering in a unified framework, highlighting the underlying relationships between them. We propose an Order Selective RLS algorithm and recursive approximations to the maximum likelihood and least p-norm estimators. Simulations show the performance of this family of algorithms for channel identification in impulsive noise environment.

1. INTRODUCTION

The optimisation criterion in Gaussian noise environment is often the minimisation of a quadratic function of the estimation error. In this paper we assume that the interfering noise exhibits impulsive behaviour and is modelled as a symmetric α-stable stochastic process. It is well known [1] that for a non-Gaussian stable distribution with characteristic exponent α, only moments of order less than α are finite, making the use of quadratic criteria as a measure of dispersion meaningless. On the other hand, the stable distribution is best used to model signals and noise that exhibit impulsive nature. This type of signals tends to produce "outliers". It is however known that, although least squares (LS) signal processing is adequate under the Gaussian assumption, it is not robust against outliers [2–4]. In this paper we propose a unified framework for a family of robust LS-type recursive algorithms for adaptive channel identification in α-stable noise environment.

The paper is organised as follows. In section 2 we present the basic assumptions of a channel identification scenario. Section 3 is a quick overview of the class of α-stable random variables. Three robust LS algorithms are presented in section 4 and the underlying relationships between them are identified under a unified framework. Section 5 presents some experimental results which show the performance of the discussed algorithms. Finally, section 6 discusses some concluding remarks.

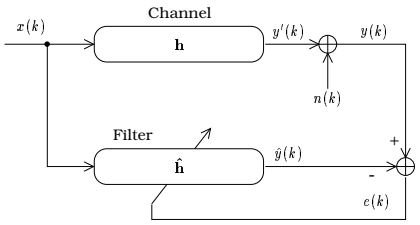


Figure 1: Channel identification model

2. SYSTEM MODEL

We assume that the digital data sequence $\{x(k) = +1, -1\}$ is passed through a noiseless linear channel with finite impulse response $\mathbf{h} = [h_0,\ h_1,\ \ldots\ h_{N-1}]^{\mathrm{T}}$ (see fig. 1). The output of the channel is

$$y'(k) = \mathbf{h}^{\mathrm{T}}\mathbf{x}(k) \tag{1}$$

The observed sequence is formed by adding random noise to the output of the channel

$$y(k) = y'(k) + n(k) \tag{2}$$

The random noise sequence $\{n(k)\}$ is modelled as a symmetric α-stable process ($S\alpha S$). Suppose that the filter operates in supervised mode. Let $\hat{\mathbf{h}}$ be the filter's coefficients vector. Its output is then

$$\hat{y}(k) = \hat{\mathbf{h}}^{\mathrm{T}}\mathbf{x}(k) \tag{3}$$

and the estimation error

$$e(k) = y(k) - \hat{y}(k) \tag{4}$$

3. THE CLASS OF α-STABLE RANDOM VARIABLES

The stable law is a direct generalisation of the Gaussian distribution and in fact includes the Gaussian as a limiting case[1]. The main difference between the non-Gaussian stable distribution and the Gaussian distribution is that the tails of the stable density are heavier than those of the Gaussian density. This characteristic of the stable distribution is one of the main reasons why the stable distribution is suitable for modelling signals and noise of impulsive nature. The $S\alpha S$ probability density function (pdf) is defined by means of its characteristic function [2]

$$F(\omega) = \exp(\delta i\omega - \gamma|\omega|^{\alpha}) \tag{5}$$

where i) α ($0 < \alpha \leq 2$) is the characteristic exponent, controlling the heaviness of the pdf tails, ii) γ ($\gamma > 0$) is the dispersion, which plays an analogous role to the variance, and iii) δ is the location parameter, the symmetry axis of the pdf. For a more detailed discussion of α-stable processes refer to [1, 5, 6].

4. ROBUST CHANNEL IDENTIFICATION

In [1] the authors introduce the *least mean p-norm* (LMP) algorithm as a direct generalisation of LMS, but it seems to share the same disadvantages with LMS. The need, however, for faster converging algorithms often calls for LS signal processing. The effort to provide robust versions of LS algorithms for non-Gaussian noise environments has received attention in the literature [2–4, 7]. In this section we discuss robust LS type algorithms for adaptive channel identification.

4.1. Order Selective Recursive Least Squares (OSRLS)

First, we propose a heuristic method to deal with the outliers of non-Gaussian distributions. The idea is to discard samples that contain large interference levels and defer the estimate update until a suitable sample is received. Specifically, in order to identify the outliers, order statistics is used. We keep a sorted vector with the magnitude of the last Θ estimation error samples. If the current error sample lies among the top η larger past samples, the current observation is discarded. Otherwise it is used as input to the recursive least squares (RLS) [8, 9] algorithm in order to update the filter coefficients. Formally, the Order Selective RLS algorithm can be expressed as follows

OSRLS(η, Θ) Algorithm
Initialisation: $\mathbf{P}(0) = \delta^{-1}\mathbf{I}, \quad \hat{\mathbf{h}}(0) = \mathbf{0}$
Basic recursion: $k = 0, 1, 2, \ldots$

1. Construct the sorted error magnitude vector

$$\bar{\mathbf{e}}(k) = \text{sort}([|e(k)|,\ |e(k-1)|, \ldots |e(k-\Theta+1)|])$$

2. Find the rank of $e(k)$ in $\bar{\mathbf{e}}(k)$

$$\mathcal{R}(k) = \text{rank}_{\bar{\mathbf{e}}(k)}(e(k))$$

3. if $(\mathcal{R}(k) < \Theta - \eta)$ *Perform the RLS recursion*
4. *repeat from step 1*

As the experimental results suggest, OSRLS can achieve good performance in highly impulsive environments. Its main disadvantage, though, is that there is no way of determining the optimal values for the parameters η and Θ.

4.2. Recursive Weighted LS (RWLS)

Huber [7] proposed the class of M-estimators as a robust version of the LS estimate. Instead of minimising the sum of squared error, Huber proposed to minimise a sum of a less rapidly increasing function ρ of the error

$$\mathcal{J}_M = \sum_{i=0}^{k} \lambda^{k-i}\rho(e(i)) \tag{6}$$

The parameter λ ($0 < \lambda \leq 1$) is the exponential weighting factor. Suppose that ρ has a derivative $\psi = \rho'$; then, the minimisation of eq. (6) implies

$$\sum_{i=0}^{k} \lambda^{k-i}\psi(e(i))x(i-j) = 0, \quad j = 0, 1, \ldots N-1 \tag{7}$$

Now, let

$$\phi(x) = \frac{\psi(x)}{x}, \quad \text{and} \quad v(i) = \phi(e(i)) \tag{8}$$

Equation (7) can be rewritten as follows

$$\sum_{i=0}^{k} \lambda^{k-i}v(i)e(i)x(i-j) = 0, \quad j = 0, 1, \ldots N-1 \tag{9}$$

Equation (9) may be viewed as a *soft-selective* version (compared to the *hard* selection of OSRLS) of the traditional LS techniques. The summation terms are not all

[1] In general, stable distributions are usually assumed to be non-Gaussian although the Gaussian distribution belongs to the family of stable distributions

[2] The characteristic function $F_{\mathbf{x}}(\omega)$ of a random variable $\mathbf{x}$ is the Fourier transform of its probability density function $f_{\mathbf{x}}(x)$

incorporated linearly in the summation, but the smaller ones are weighted more heavily, depending on the specific penalty function ρ. The actual weighting is controlled by the *weighting function* ϕ. On the other hand, eq. (9) can be rewritten as a generalisation of the LS equations

$$\sum_{i=0}^{k} \lambda^{k-i} e_{\mathrm{g}}(i) x_{\mathrm{g}}(i-j) = 0, \quad j = 0, 1, \ldots N-1 \quad (10)$$

where

$$\begin{aligned} x_{\mathrm{g}}(i-j) &= v^{\frac{1}{2}}(i) x(i-j) \\ y_{\mathrm{g}}(i) &= v^{\frac{1}{2}}(i) y(i) \\ e_{\mathrm{g}}(i) &= v^{\frac{1}{2}}(i) e(i) \end{aligned} \quad (11)$$

Equation (10) is essentially a *generalised orthogonality criterion*. This generalisation implies (as in [10]) that all quadratic adaptive filter algorithms can be made robust if the vector $\mathbf{x}(i)$ and the target $y(i)$ at time i are modified according to eq. (11). In fact, LMP [1] can be derived from LMS using this generalisation.

After some rearrangements and exchange in summation order in eq. (9) we may express this set of equations in matrix form

$$\left[\sum_{i=0}^{k} \lambda^{k-i} v(i) \mathbf{x}(i) \mathbf{x}^{\mathrm{T}}(i)\right] \hat{\mathbf{h}}(k) = \left[\sum_{i=0}^{k} \lambda^{k-i} v(i) \mathbf{x}(i) y(i)\right] \quad (12)$$

The sequence $v(i)$ assumes knowledge of the optimal weight vector $\hat{\mathbf{h}}$ at time k to generate the error sequence $e(i)$. As in [2, 11], for the recursive approximation of the tap weight estimate $\hat{\mathbf{h}}$ the instantaneous *a priori* estimation error $\xi(i) = y(i) - \hat{\mathbf{h}}^{\mathrm{T}}(i-1)\mathbf{x}(i)$ is used to approximate $e(i)$. We can, therefore, generate the sequence $w(i) = \phi(\xi(i))$ in order to approximate $v(i)$. We now define the *weighted autocorrelation matrix* of the data $\mathbf{Q}_{xx}$ and the *weighted cross-correlation vector* $\mathbf{q}_{xy}$ as [3, 7]

$$\mathbf{Q}_{xx}(k) = \lambda \mathbf{Q}_{xx}(k-1) + w(k)\mathbf{x}(k)\mathbf{x}^{\mathrm{T}}(k) \quad (13)$$

$$\mathbf{q}_{xy}(k) = \lambda \mathbf{q}_{xy}(k-1) + w(k)\mathbf{x}(k)y(k) \quad (14)$$

Using these definitions, we may use the following approximation to eq. (12)

$$\mathbf{Q}_{xx}(k)\hat{\mathbf{h}}(k) = \mathbf{q}_{xy}(k) \quad (15)$$

Applying the *matrix inversion lemma* [8, 9] to eq. (14) yields the *Recursive Weighted Least Squares* algorithm which can be summarised

RWLS Algorithm
Initialisation: $\mathbf{P}(0) = \delta^{-1}\mathbf{I}, \quad \hat{\mathbf{h}}(0) = \mathbf{0}$
Basic recursion: $k = 0, 1, 2, \ldots$

1. Gain vector

$$\mathbf{K}(k) = \frac{\lambda^{-1}\mathbf{P}(k-1)\mathbf{x}(k)}{w(k)^{-1} + \lambda^{-1}\mathbf{x}^{\mathrm{T}}(k)\mathbf{P}(k-1)\mathbf{x}(k)}$$

2. Tap weight vector update

$$\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k-1) + \xi(k)\mathbf{K}(k)$$

3. Inverse weighted autocorrelation matrix

$$\mathbf{P}(k) = \lambda^{-1}\mathbf{P}(k-1) - \lambda^{-1}\mathbf{K}(k)\mathbf{x}^{\mathrm{T}}(k)\mathbf{P}(k-1)$$

4. *repeat from step 1*

where

$$\mathbf{P}(k) = \mathbf{Q}_{xx}^{-1}(k) \quad (16)$$

$$\mathbf{K}(k) = w(k)\mathbf{P}(k)\mathbf{x}(k) \quad (17)$$

It is interesting to notice that the recursive equations of RWLS are almost identical to the traditional recursive least squares (RLS) algorithm. On the other hand, the same set of recursive equations can be derived from RLS using the generalised orthogonality criterion (eq. (10)). Moreover, eq. (17) implies

$$\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k-1) + \mathbf{P}(k)\mathbf{x}(k)\psi(\xi(k)) \quad (18)$$

This is essentially equivalent to the recursive Gauss-Newton algorithm for general criteria [12] which can be summarised as follows [3]

$$\mathbf{R}(k) = \mathbf{R}(k-1) + \gamma(k)[\mathbf{x}(k)\rho''(\xi(k))\mathbf{x}^{\mathrm{T}}(k) - \mathbf{R}(k-1)] \quad (19)$$

$$\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k-1) + \gamma(k)\mathbf{R}^{-1}(k)\mathbf{x}(k)\rho'(\xi(k)) \quad (20)$$

4.2.1. Traditional Recursive Least Squares

The RWLS recursion is reduced to the traditional RLS algorithm if a quadratic penalty function is chosen, corresponding to

$$\begin{aligned} \rho_{LS}(x) &= \frac{x^2}{2} \\ \psi_{LS}(x) &= x \\ \phi_{LS}(x) &= 1 \end{aligned} \quad (21)$$

[3] Note that for a polynomial penalty function $\rho(x)$, its second derivative $\rho''(x)$ is proportional to $\phi(x)$ (see eq. (8))

4.2.2. Recursive Maximum Likelihood (RML)

Assume that the noise samples are independent identically distributed (i.i.d.), drawn from a random variable which pdf is f. Then, the likelihood function of the received vector $\mathbf{y}$ under the parameters $\hat{\mathbf{h}}$ is given by

$$\mathcal{L}_{\hat{\mathbf{h}}}(\mathbf{y}; f) \triangleq \log \prod_{i=0}^{k} f(e(i)) = \sum_{i=0}^{k} \log f(e(i)) \qquad (22)$$

Therefore, the maximum likelihood estimate (ML) $\hat{\mathbf{h}}$ is obtained by RWLS when

$$\begin{aligned} \rho_{ML}(x) &= -\log f(x) \\ \psi_{ML}(x) &= -\frac{f'(x)}{f(x)} \\ \phi_{ML}(x) &= -\frac{f'(x)}{x f(x)} \end{aligned} \qquad (23)$$

The RML, however, is not of practical interest, as it requires that the noise pdf is known to the receiver. In general, there is no closed-form solution for the $S\alpha S$ pdf. For our simulations, in the following section, numerical approximations of the $S\alpha S$ distribution have been used.

4.2.3. Recursive Least p-norm

As shown in [1], the *minimum dispersion* criterion is a natural and mathematically meaningful choice as a measure of optimality in stable signal processing. By minimising the error dispersion, we minimise the average magnitude of the estimation error. Furthermore, it has been shown that minimising the dispersion is also equivalent to minimising the probability of large estimation errors [13]. The norm of a $S\alpha S$ random variable is proportional to its p-th order moment for any $0 < p \leq \alpha$. Consequently, the appropriate cost function would be

$$\mathcal{J}_{LP} = \sum_{i=0}^{k} |e(i)|^p \qquad (24)$$

and, therefore

$$\begin{aligned} \rho_{LP}(x) &= \frac{|x|^p}{p} \\ \psi_{LP}(x) &= \langle x \rangle^{p-1} \\ \phi_{LP}(x) &= |x|^{p-2} \end{aligned} \qquad (25)$$

where

$$\langle x \rangle^p = \text{sign}(x)\, |x|^p$$

The choice of the weighting sequence $w(i)$ from eq. (25) for RWLS gives rise to the *recursive least p-norm* (RLP) algorithm for adaptive filtering in $S\alpha S$ noise environment. According to eq. (25), one does not have to explicitly estimate the actual noise pdf, but only set parameter p to be less than α. Equation (25) also suggests that $w(i)$ is, in general, not bounded. That is, for infinitesimally small error the corresponding weight is infinitely large. If we consider eq. (13) and eq. (14), infinitely large terms may result to numerical instability in the computation of $\mathbf{Q}_{xx}$ and $\mathbf{q}_{xy}$. Moreover, theoretical justifications [7, 12, 14] require a bounded weighting sequence. We can bound the sequence $w(i)$ for very small values of and normalise

$$w(i) = \begin{cases} |\frac{\xi(i)}{\omega}|^{p-2} & , \quad |\xi(i)| > \omega \\ 1 & , \quad |\xi(i)| \leq \omega \end{cases} \qquad (26)$$

where ω is a small positive number. Essentially, for samples with small estimation error (where probably the noise term is very small) we use the traditional LS penalty function ($\rho(x) = x^2/2$).

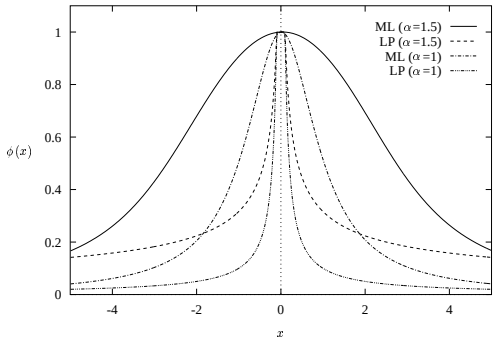


Figure 2: Weighting functions for Maximum Likelihood and Least p-norm criteria for two values of the characteristic exponent

5. SIMULATION EXAMPLES

In fig. 2 we plot the weighting functions ϕ for the two non-linear optimisation criteria (ML and LP), for the cases $\alpha = 1.5$ and $\alpha = 1$. Note that the larger the magnitude of the residuals, the smaller the weight both algorithms produce. Indeed, large residuals are likely to be the result of noise impulses.

Experiments for channel identification using the algorithms LMP, OSRLS, RML and RLP have been carried out with channel impulse response

$$\begin{aligned} \mathbf{h}_1 = [&.04, \quad -.05, \quad .07, \quad -.21, \quad -.5, \quad .72, \\ &.36, \quad 0, \quad .21, \quad .03, \quad .07]^{\mathrm{T}} \end{aligned}$$

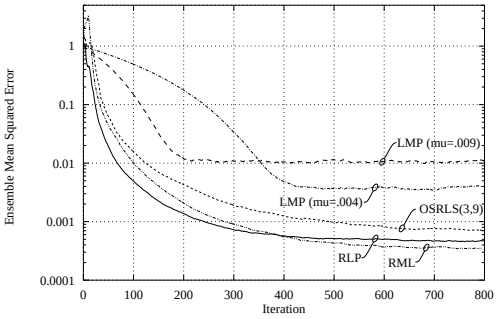


Figure 3: The convergence of LMP, OSRLS, RML and RLP for channel with impulse response $\mathbf{h}_1$

and noise parameters $\alpha = 1$, $\gamma = .08$, $\delta = 0$. The step-size parameter for LMP was $\mu = .009$ and $\mu = .004$. The forgetting factor λ was set to 0.98. For OSRLS, $\eta = 3$ and $\Theta = 9$. For RLP, ω was chosen 0.1. Numerical approximation of the weighting function ϕ was used for RML.

Figure 3 depicts the ensemble (over 400 runs) mean squared error for algorithms LMP, OSRLS, RML and RLP. Clearly all LS type algorithms outperform the stochastic gradient LMP. The performance of OSRLS is poorer than both RML and RLP because this algorithm discards about a third of the received samples. The best asymptotic performance is achieved, as expected, by RML, but the performance of RLP is very close. In fact, RLP converges faster than RML in the first 400 iterations.

6. CONCLUDING REMARKS

In this paper we propose a unified framework for robust adaptive LS type filtering in $S\alpha S$ impulsive noise environment and present three new algorithms. According to experimental results, RML achieves the best asymptotic performance but its use in practice is prohibited by its computational load. On the other hand RLP offers less computational complexity at the cost of a small degradation in convergence. The algorithms are presented in a channel identification context. However, their use could be also beneficial in adaptive linear equalisation, in training the centers of radial basis functions (RBF) networks and multi-user detection in non-Gaussian channels.

7. REFERENCES

[1] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proceedings of the IEEE*, vol. 81, pp. 986–1009, July 1993.

[2] H. Dai and N.K. Sinha, "Robust recursive least-squares method with modified weights for bilinear system identification," *Proceedings of the IEE*, vol. 136, no. 3, pp. 122–126, May 1989.

[3] E.Kuruoğlu, W.Fitzgerald, and P.Rayner, "Non-linear autoregressive modeling of non-Gaussian signals using L_p-norm techniques," *International Conference on Acoustics, Speech and Signal Processing*, vol. 3, pp. 3533–3536, 1997.

[4] X. Wang and H.V. Poor, "Robust multi-user detection in non-Gaussian channels," *IEEE Transactions on Signal Processing*, vol. 47, no. 2, pp. 289–305, February 1999.

[5] G.Samorodnitsky and M.S.Taqqu, *Stable non-Gaussian random processes*, Chapman & Hall, 1994.

[6] D. Middleton, "Non-Gaussian noise models in signal processing for telecommunications: New methods and results for class A and class B noise models," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1129–1149, May 1999.

[7] P.J. Huber, *Robust Statistics*, Wiley, New York, 1981.

[8] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 3rd edition, 1997.

[9] B. Mulgrew and C.F.N. Cowan, *Adaptive filters and equalisers*, Kluwer Academic Publishers, 1988.

[10] J. Bodenschatz and C. Nikias, "Recursive Local Orthogonality Filtering," *IEEE Transactions on Signal Processing*, vol. 45, no. 9, pp. 2293–2300, September 1997.

[11] S.C. Puthenpura, N.K. Sinha, and O.P. Vidal, "Application of M-estimation in robust recursive system identification," *IFAC Symp. Stocastic Control*, pp. 23–30, 1985.

[12] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*, MIT Press, 1983.

[13] D.B.Cline and P.J.Brockwell, "Linear prediction of ARMA processes with infinite variance," *Stochastic Processes and Th*, vol. 19, pp. 281–296, 1985.

[14] R.Yarlagadda, J.B.Bednar, and T.Watt, "Fast algorithms for l_p deconvolution," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, pp. 174–182, 1985.

Adaptive Bayesian Decision Feedback Equaliser for Alpha–Stable Noise Environments

Apostolos T. Georgiadis and Bernard Mulgrew

Department of Electronics & Electrical Engineering
University of Edinburgh, EH9 3JL Edinburgh, UK
e-mail: atg@ee.ed.ac.uk

Abstract

In some communication systems the channel noise is known to be non–Gaussian due, largely, to impulsive phenomena. The performance of signal processing algorithms designed under the Gaussian assumption may degrade seriously in such environments. In this paper we investigate the problem of adaptive channel equalisation in an impulsive noise environment. The impulsive interfering noise is modelled as an α–stable process. We first derive the optimum Bayesian decision feedback equaliser and present a novel analytical framework for the evaluation of systems in infinite variance environments. A family of generalised adaptive channel identification algorithms for this infinite variance noise environment is also presented. The combination of a Bayesian equaliser and a channel estimator operating as an adaptive channel equaliser is experimentally studied and its performance is compared with that of a traditional system designed under the Gaussian assumption. The experimental data suggest that the proposed combination of equaliser and channel estimator outperforms the traditionally designed adaptive equaliser in terms of error probability. We finally provide some useful approximations concerning the practical implementation of an α–stable adaptive equaliser.

1 Introduction

High speed data transmission over communication channels is subject to intersymbol interference and noise. The intersymbol interference is usually the result of the restricted bandwidth allocated to the channel and/or the presence of multipath distortion in the medium through which the information is transmitted. Equalisation is the process which reconstructs the transmitted data combating the distortion and interference of the communication link. The most simple architecture in the class of equalisers making decisions in a

Preprint submitted to Elsevier Preprint *21 November 2000*

symbol–by–symbol basis is the linear transversal filter. The optimal solution, however, is the Bayesian approach which is also known as the maximum a posteriori (MAP) symbol–by– symbol decision equaliser [1].

Although the Bayesian equaliser and its adaptive implementation has been thoroughly studied in the literature (for example see [2] and the references therein), by and large, the results are related to the assumption that the interference noise is Gaussian. However, in many physical channels, such as urban, indoor radio and underwater acoustic channels [3–5], the ambient noise is known through experimental measurements to be non-Gaussian, mainly due to the impulsive nature of man–made electromagnetic interference. It is well known that non–Gaussian noise can cause significant performance degradation in conventional systems based on the Gaussian assumption [5].

A number of models have been proposed for impulsive phenomena in communication systems, either by fitting experimental data or based on physical grounds. Recently, it has been suggested [5] that the family of α–stable random variables provides an appropriate model for many impulsive phenomena, including interference in communication channels. Stable distributions share defining characteristics with the Gaussian distribution, such as the stability property and central limit theorems.

In the following, after a quick overview of stable processes (section 2), we derive in section 3 the optimum Bayesian decision feedback equaliser (DFE) for α–stable noise environments. The problem of evaluating communication systems in infinite variance environments is addressed in section 4 and a new analytical framework in this direction is presented. Some preliminary experimental results are given in section 4.1, showing a promising performance benefit compared with a Bayesian DFE designed under the Gaussian assumption [1]. Section 5 discusses the problem of estimating the channel and noise characteristics in an α–stable noise environment. A family of recursive algorithms for channel identification in such environments is presented and studied. The adaptive Bayesian DFE is then experimentally studied in section 6. Some useful approximations concerning the simulation and implementation of such an equaliser are finally discussed in section 7.

2 The class of stable random variables

The family of stable random variables (RV) is defined as a direct generalisation of the Gaussian law. The main characteristic of a non–Gaussian stable

[1] Hereinafter, a Bayesian equaliser designed under the Gaussian assumption will be referred to as *traditional Bayesian equaliser*.

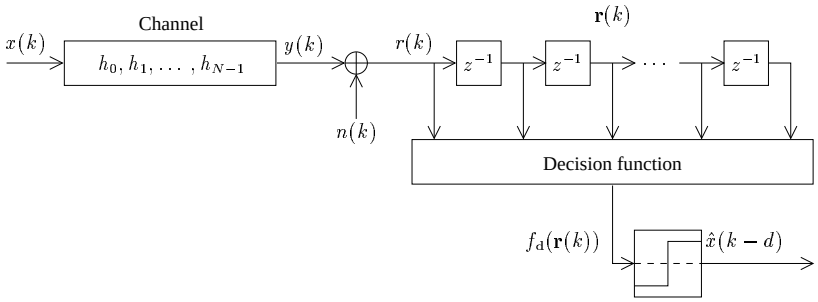


Fig. 1. System model for FIR channel and finite memory equaliser

probability density function (pdf) is that its tails are heavier than those of the normal density. This is one of the main reasons why the stable law is regarded suitable for modelling signals and noise of impulsive nature.

The symmetric α–stable ($S\alpha S$) pdf $f_\alpha(s)$ is defined by means of its characteristic function [2] $F(\omega) = \exp(\delta i\omega - \gamma|\omega|^\alpha)$. The parameters α, γ and δ describe completely a $S\alpha S$ distribution. The characteristic exponent α ($0 < \alpha \leq 2$) controls the heaviness of the tails of the stable density; a smaller value implies heavier tails, while $\alpha = 2$ is the Gaussian case. The dispersion parameter γ ($\gamma > 0$) plays an analogous role to the variance and refers to the spread of the distribution. Finally, the location parameter δ is comparable with the mean of the distribution. In fact they are identical for $1 < \alpha \leq 2$.

Theoretical justifications for using the stable distribution as a basic statistical modelling tool come from the generalised central limit theorem [6].Unfortunately, no closed–form expressions exist for the stable density, except the Gaussian ($\alpha = 2$) and Cauchy ($\alpha = 1$) distributions. An important property of all stable distributions is that only the lower order moments are finite. That is, if $\mathbf{x}$ is a stable RV, then $\boldsymbol{E}_{\mathbf{x}}\{|x|^p\} < \infty$ iff $p < \alpha$. A well known consequence of this property is that all stable RV's with $\alpha < 2$ have infinite variance. For a more detailed discussion of α–stable processes refer to [7]. Moreover, [5] presents a signal processing framework for α–stable processes.

3 Bayesian equaliser

The model of the system considered is depicted in fig. 1. We assume that the data sequence $\{x(k) = +1, -1\}$, consisting of independent and equiprobable binary symbols, is passed through a noiseless linear dispersive channel with

[2] The characteristic function $F(\omega)$ of a RV is the Fourier transform of its probability density function $f(s)$.

finite impulse response (FIR) which spans over N symbols:

$$H(z) = \sum_{i=0}^{N-1} h_i z^{-i}, \quad \mathbf{h} = [h_0, h_1 \dots h_{N-1}]^{\mathrm{T}} \tag{1}$$

If $\mathbf{x}_{\mathrm{ch}}(k) = [x(k)\ x(k-1)\ \dots\ x(k-N+1)]^{\mathrm{T}}$ is the channel input vector, then the observation sequence $\{r(k)\}$ is formed by adding the α–stable random noise $n(k)$ to the output of the channel $y(k) = \mathbf{h}^{\mathrm{T}}\mathbf{x}_{\mathrm{ch}}(k)$, i.e., $r(k) = y(k) + n(k)$. In finite memory equalisers, the M most recent samples of the observation sequence $\{r(k)\}$ are stored in the observation vector

$$\mathbf{r}(k) = [r(k) \quad r(k-1)\ \cdots\ r(k-M+1)]^{\mathrm{T}} \tag{2}$$

A *decision function* $f_{\mathrm{d}}(\cdot)$ is then evaluated on $\mathbf{r}(k)$ and passed through a quantiser to provide an estimate of the transmitted symbol $x(k-d)$. Here, d is the decision lag of the equaliser.

3.1 Feed-forward equaliser

Let $\mathbf{x}(k)$ be the vector with all the transmitted symbols that influence $\mathbf{r}(k)$, i.e.,

$$\mathbf{x}(k) = [x(k) \quad x(k-1)\ \cdots\ x(k-K+1)]^{\mathrm{T}} \tag{3}$$

where $K = N + M - 1$. The state equation that relates the received vector $\mathbf{r}(k)$ to $\mathbf{x}(k)$ is

$$\mathbf{r}(k) = \mathbf{H} \cdot \mathbf{x}(k) + \mathbf{n}(k) \tag{4}$$

where

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & h_2 & \cdots & h_{N-1} & 0 & 0 \\ 0 & h_0 & h_1 & \cdots & h_{N-2} & h_{N-1} & 0 \\ & \vdots & & & & \ddots & \\ 0 & 0 & 0 & \cdots & h_0 & \cdots & h_{N-1} \end{bmatrix} \tag{5}$$

is the $M \times K$ channel matrix, and $\mathbf{n}(k)$ contains the noise samples

$$\mathbf{n}(k) = [n(k) \quad n(k-1)\ \cdots\ n(k-M+1)]^{\mathrm{T}} \tag{6}$$

There are totally $N_c = 2^K$ possible discrete states [3] $\mathbf{c}_i$ for the noise–free observation vector $\mathbf{y}(k) = \mathbf{H} \cdot \mathbf{x}(k)$. These states $\mathbf{c}_i$ can be partitioned into

[3] Also referred as *centers*.

two sets, conditioned on the transmitted symbol of interest [2]

$$S^+ = [\mathbf{c}_i \mid x(k-d) = +1] \quad \text{, and} \quad S^- = [\mathbf{c}_i \mid x(k-d) = -1]$$

As in [8], the appropriate MAP decision function is

$$\begin{aligned} f_{\mathrm{d}}(\mathbf{r}(k)) &= \sum_{\mathbf{c}_i} s_i p_{\mathbf{r} \mid \mathbf{c}}(\mathbf{r}(k)|\mathbf{c}_i) P(\mathbf{c}_i) \\ &= \frac{1}{N_c} \sum_{\mathbf{c}_i} s_i f_\alpha(\mathbf{r}_0(k)|\mathbf{c}_{i,0}) f_\alpha(\mathbf{r}_1(k)|\mathbf{c}_{i,1}) \cdots f_\alpha(\mathbf{r}_{M-1}(k)|\mathbf{c}_{i,M-1}) \end{aligned} \quad (7)$$

where $p_{\mathbf{r} \mid \mathbf{c}}$ is the likelihood of $\mathbf{r}$ conditioned on $\mathbf{c}$, $P(\mathbf{c}_i)$ is the *a priori* probability that $\mathbf{c}_i$ occurs, $f_\alpha(s)$ is the pdf of the $S\alpha S$ additive noise, and

$$\mathbf{r}_i(k) = r(k-i) \quad , \quad s_i = \begin{cases} +1 & : \quad \mathbf{c}_i \, \epsilon \, S^+ \\ -1 & : \quad \mathbf{c}_i \, \epsilon \, S^- \end{cases}$$

Note that eq. (7) reduces to the traditional MAP equaliser for $\alpha = 2$. The actual estimate is given by the sign of f_{d}, i.e.,

$$\hat{x}(k-d) = \mathrm{sign}\left(f_{\mathrm{d}}(\mathbf{r}(k))\right) \quad (8)$$

Equation (8) partitions the M–dimensional observation space spanned by the received signal vector $\mathbf{r}(k)$ in two sub–spaces. Therefore, the solution of equation $f_{\mathrm{d}}(\mathbf{r}(k)) = 0$ defines the optimum decision boundary. Since $f_{\mathrm{d}}(\mathbf{r}(k))$ is related to the pdf of the noise, the corresponding Bayesian decision boundaries will be inherently different for Gaussian and non–Gaussian distributions. In [2] a radial basis functions network implementation of eq. (7) is suggested. However, it has been demonstrated [8] that in non–Gaussian noise environments the basis functions are not radially symmetric. The actual radial asymmetry of the M–dimensional stable noise pdf is responsible for the radical discrepancies between the Gaussian and non-Gaussian decision boundaries.

3.2 Decision feedback equaliser

Without loss of generality we can assume that the D decisions $\hat{x}(k-L)$, $\hat{x}(k-L-1) \ldots \hat{x}(k-K+1)$ are correct (here $L = K - D$). Replacing these decisions [9–11] on the trailing part of vector $\mathbf{x}(k)$ we have

$$\hat{\mathbf{x}}(k) = [x(k) \quad \ldots \; x(k-L+1) \,\big|\, \hat{x}(k-L) \; \ldots \; \hat{x}(k-K+1)]^{\mathrm{T}}$$

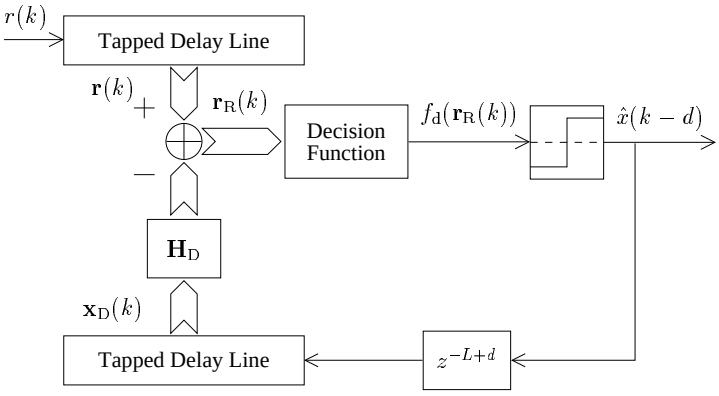


Fig. 2. Decision feedback equaliser

If we now appropriately partition the channel matrix as

$$\mathbf{H} = \left[\begin{array}{cccc|ccccc} h_0 & h_1 & \cdots & h_{L-1} & h_L & \cdots & h_{N-1} & 0 & \cdots & 0 \\ 0 & h_0 & \cdots & h_L & h_{L+1} & \cdots & h_{N-2} & h_{N-1} & \cdots & 0 \\ & \vdots & & & & & \ddots & & & \\ 0 & 0 & \cdots & & & & & \cdots & & h_{N-1} \end{array}\right] \tag{9}$$

we can rewrite eq. (4) as [2]

$$\mathbf{r}(k) = \left[\mathbf{H}_{\mathrm{R}} \middle| \mathbf{H}_{\mathrm{D}}\right] \left[\frac{\mathbf{x}_{\mathrm{R}}(k)}{\mathbf{x}_{\mathrm{D}}(k)}\right] + \mathbf{n}(k) \tag{10}$$

The sub–matrices $\mathbf{H}_{\mathrm{R}}$, $\mathbf{H}_{\mathrm{D}}$, $\mathbf{x}_{\mathrm{R}}$, and $\mathbf{x}_{\mathrm{D}}$ are defined in an obvious manner [4].

The effect of the decisions contained in $\mathbf{x}_{\mathrm{D}}(k)$ can then be removed from the observation vector $\mathbf{r}(k)$ to produce a *residual* observation vector, defined as

$$\mathbf{r}_{\mathrm{R}}(k) \triangleq \mathbf{r}(k) - \mathbf{H}_{\mathrm{D}}\mathbf{x}_{\mathrm{D}}(k) = \mathbf{H}_{\mathrm{R}}\mathbf{x}_{\mathrm{R}}(k) + \mathbf{n}(k) = \mathbf{y}_{\mathrm{R}}(k) + \mathbf{n}(k) \tag{11}$$

We can now apply a Bayesian decision function to $\mathbf{r}_{\mathrm{R}}(k)$ rather than $\mathbf{r}(k)$.

A decision feedback equaliser implementing this scheme is depicted in fig. 2. In fig. 3 we can see the optimum boundaries for the Bayesian DFE with feedback order $D = 2$ and a variety of values for the characteristic exponent. The features of the optimum decision boundaries are significantly different compared to the boundaries of a traditional MAP equaliser. Therefore, it is reasonable to expect a considerable performance degradation of the traditional Bayesian equaliser in a non–Gaussian noise environment.

[4] The subscript R stands for *residual* while D stands for *feedback*.

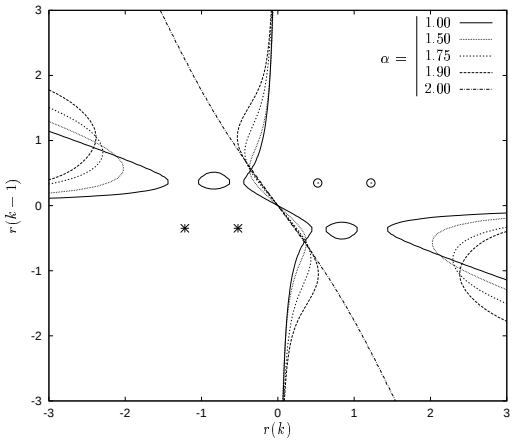


Fig. 3. Observation space and decision boundaries of eq. (7) for the Bayesian DFE. The channel is $H(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$ (the stars ✳ indicate a center in the S^- subset and the circles ⊙ a center in the S^+ subset)

4 Evaluating systems in infinite power noise environment

The traditional performance measures are usually plots of the bit–error ratio (BER) against the signal–to–noise ratio (SNR). In non–Gaussian stable noise environment (α–stable noise with $\alpha < 2$), however, the variance of the noise is infinite [5], making the use of SNR meaningless. Nevertheless, all receivers in practice have a finite input dynamic range. Let us consider the generic receiver depicted in fig. 4. The limiter at the front end of the receiver is assumed to be an ideal saturation device, with transfer function

$$g(s, G) = \begin{cases} s & : \quad |s| \leq G \\ \text{sign}(s)\ G & : \quad \text{elsewhere} \end{cases} \tag{12}$$

G being the saturation point of the limiter. For a given saturation limit G, the SNR for the limited received signal $r_{\text{L}}(k)$ is always finite. In this paper, we propose that the SNR at the limited received signal $r_{\text{L}}(k)$ should be used for performance evaluation in environments where the noise variance is infinite. We will refer to this as the *SNR at the receiver*. In the following we present some analytical tools that enable us to calculate this SNR.

The distribution of the received signal $r(k)$ is

$$f_r(s) = \frac{1}{N_{\text{sc}}} \sum_{i=1}^{N_{\text{sc}}} f_\alpha(s - \bar{c}_i) \tag{13}$$

where $N_{\text{sc}} = 2^N$ is the number of the scalar centers [5] $\bar{c}_i$ of the channel, i.e.,

[5] Scalar centers are all the discrete noise–free channel outputs.

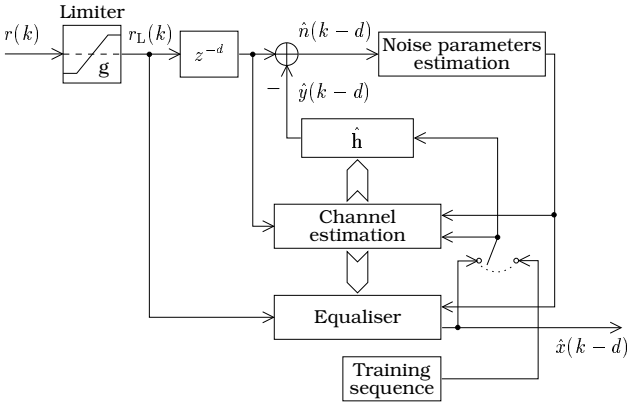


Fig. 4. Generic adaptive equaliser with saturation device at the front end

$\bar{c}_i = \mathbf{h}^{\mathrm{T}} \cdot \mathbf{x}_{\mathrm{ch_i}}$ $(i = 1, 2, \ldots\ N_{\mathrm{sc}})$. Here $\mathbf{x}_{\mathrm{ch_i}}$ are all the possible combinations for the channel input vector. The limiter g truncates the pdf of the received signal and its tails are concentrated at the points $+G, -G$ where they appear as *Dirac* impulses $\delta(s)$ (fig. 5(a)). The pdf of the limited received sequence $r_{\mathrm{L}}(k)$ is therefore

$$f_{r_{\mathrm{L}}}(s) = \frac{1}{N_{\mathrm{sc}}} \sum_{i=1}^{N_{\mathrm{sc}}} \bar{f}_\alpha(s - \bar{c}_i, -G - \bar{c}_i, G - \bar{c}_i) \tag{14}$$

where $\bar{f}_\alpha(s, G_1, G_2)$ is the α–stable pdf *truncated* at the points G_1 and G_2 and is given by

$$\bar{f}_\alpha(s, G_1, G_2) = f_\alpha(s)\,\Pi(s, G_1, G_2) + I_{\mathrm{l}}(G_1)\delta(s - G_1) + I_{\mathrm{r}}(G_2)\delta(s - G_2)$$

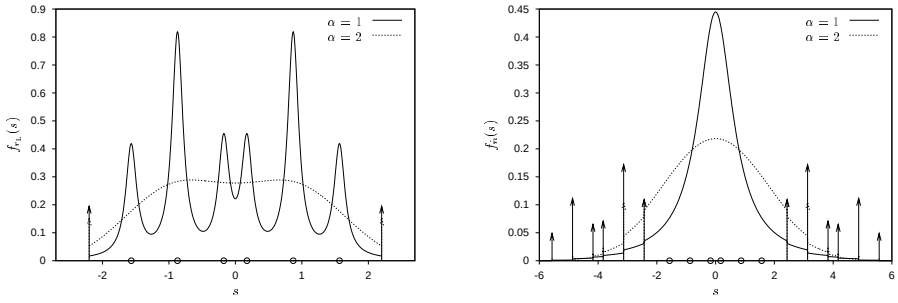
where $\Pi(s, G_1, G_2)$ is 1 within $[G1, G2]$ and 0 elsewhere, and

$$I_{\mathrm{l}}(G) = \int_{-\infty}^{G} f_\alpha(s)\,ds\ , \quad I_{\mathrm{r}}(G) = \int_{G}^{\infty} f_\alpha(s)\,ds \tag{15}$$

The receiver removes the channel output estimate $\hat{y}(k - d)$ from the limited received signal $r_{\mathrm{L}}(k - d)$ to form an estimate of the noise samples $\hat{n}(k - d)$ (fig. 4). We can assume, without loss of generality, that the samples $\hat{y}(k)$ are correct. The pdf of the noise estimate $\hat{n}(k)$ will then be

$$f_{\hat{n}}(s) = \frac{1}{N_{\mathrm{sc}}} \sum_{i=1}^{N_{\mathrm{sc}}} \bar{f}_\alpha(s, -G - \bar{c}_i, G - \bar{c}_i) \tag{16}$$

In fig. 5(b) we can see an example for the pdf of the noise estimate sequence $\hat{n}(k)$. Due to the symmetry of scalar centers, $f_{\hat{n}}(s)$ is symmetric. Therefore,



(a) For Gaussian case $\gamma = 0.135$, and for α–stable case $\gamma = 0.1$. $G = 2.2$.

(b) For Gaussian case $\gamma = 1.67$, and for α–stable case $\gamma = 0.72$. $G = 4$.

Fig. 5. The pdf of a) $r_{\mathrm{L}}(k)$, and b) $\hat{n}(k)$ for Gaussian ($\alpha = 2$) and α–stable noise ($\alpha = 1$). The channel is $H(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$ (the circles $\circ$ denote the corresponding scalar centers).

the mean of $\hat{n}(k)$ is zero, while its variance can be written

$$v_{\hat{n}}(\alpha, \gamma, G) = \int_{-\infty}^{\infty} s^2 \, f_{\hat{n}}(s) \, ds = \frac{1}{N_{\mathrm{sc}}} \sum_{i=1}^{N_{\mathrm{sc}}} \int_{-\infty}^{\infty} s^2 \, \bar{f}_{\alpha}(s, -G - \bar{c}_i, G - \bar{c}_i) ds \quad (17)$$

The integral at the rightmost part of eq. (17) can further be expressed as

$$\begin{aligned} V(\alpha, \gamma, G_1, G_2) &= \int_{-\infty}^{\infty} s^2 \, \bar{f}_{\alpha}(s, G_1, G_2) ds \\ &= \int_{-\infty}^{\infty} s^2 \, \Big\{ f_{\alpha}(s) \, \Pi(s, G_1, G_2) + I_{\mathrm{l}}(G_1) \delta(s - G_1) + I_{\mathrm{r}}(G_2) \delta(s - G_2) \Big\} \, ds \\ &= G_1^2 \, I_{\mathrm{l}}(G_1) + G_2^2 \, I_{\mathrm{r}}(G_2) + \int_{G1}^{G2} s^2 \, f_{\alpha}(s) ds \end{aligned} \quad (18)$$

In general, $f_{\alpha}(s)$ cannot be expressed in closed form except the $\alpha = 2$ and $\alpha = 1$ cases. For these two special cases, it is possible to calculate $V(\alpha, \gamma, G_1, G_2)$. For the Gaussian case ($\alpha = 2$) we obtain

$$\begin{aligned} V(2, \gamma, G_1, G_2) = \frac{G_1^2 + G_2^2}{2} &+ \frac{1}{2} (2\gamma - G_2^2) \operatorname{erf}\left(\frac{G_2}{2\sqrt{\gamma}} \right) \\ &- \frac{1}{2} (2\gamma - G_1^2) \operatorname{erf}\left(\frac{G_1}{2\sqrt{\gamma}} \right) \\ &- 2\sqrt{\frac{\gamma}{\pi}} \left(G_2 \exp\left(-\frac{G_2^2}{4\gamma} \right) + G_1 \exp\left(-\frac{G_1^2}{4\gamma} \right) \right) \end{aligned} \quad (19)$$

and for the Cauchy case ($\alpha = 1$)

$$V(1,\gamma,G_1,G_2) = \frac{G_1^2+G_2^2}{2} + \frac{\gamma\,(G_2-G_1)}{\pi} - \frac{(\gamma^2+G_2^2)\text{atan}\left(\frac{G_2}{\gamma}\right) - (\gamma^2+G_1^2)\text{atan}\left(\frac{G_1}{\gamma}\right)}{\pi} \tag{20}$$

From eq. (17) we can write the variance of the noise estimate $\hat{n}(k)$ as

$$v_{\hat{n}}(\alpha,\gamma,G) = \frac{1}{N_{\text{sc}}}\sum_{i=1}^{N_{\text{sc}}} V(\alpha,\gamma,-G-\bar{c}_i,G-\bar{c}_i) \tag{21}$$

We can now express the SNR *at the receiver* (in dB) as a function of the noise parameters α,γ and the dynamic range of the receiver G

$$\text{SNR}_{\text{rcv}} = 10\,\log\left(\frac{v_y}{v_{\hat{n}}(\alpha,\gamma,G)}\right) \tag{22}$$

where v_y is the variance of the noise–free channel output.

In practice for a given SNR_{rcv}, characteristic exponent α and dynamic range G, it is possible to numerically solve eq. (22) for the noise dispersion γ. For the values of α that it is not possible to analytically compute eq. (22) the variance of the noise estimate $\hat{n}(k)$ may be experimentally measured in order to compute the working SNR. However, in section 7 we suggest an approximate method to compute the variance $v_{\hat{n}}(\alpha,\gamma,G)$ for a given dispersion γ. Accordingly, using an analogous approximation we can obtain γ for a given SNR_{rcv}.

4.1 Experiments

In order to assess the Bayesian equaliser in an α–stable noise environment, the experimental performance of a number of feed–forward and DF equalisers was recorded. The simulations were performed for a channel with transfer function

$$H(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2} \tag{23}$$

For the moment we assume that the equaliser has perfect knowledge of the channel model and the noise characteristics. The dynamic range of the receiver is $G = 4$.

For the first set of experiments we simulated the feed–forward MAP equaliser in varying noise environments ($\alpha = 1, 1.5, 2$). The length of the observation vector was $M = 2$ and the equalisers operated with a decision lag $d = 1$.

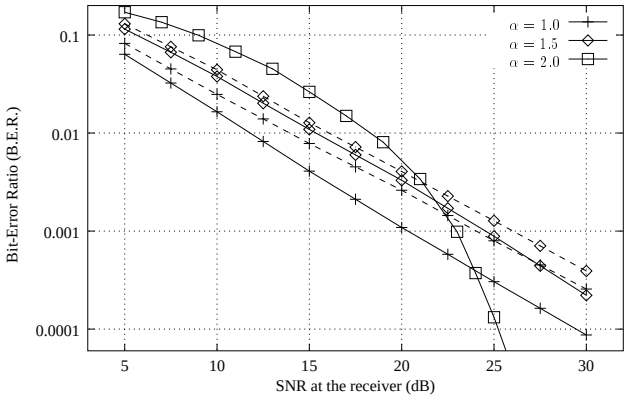


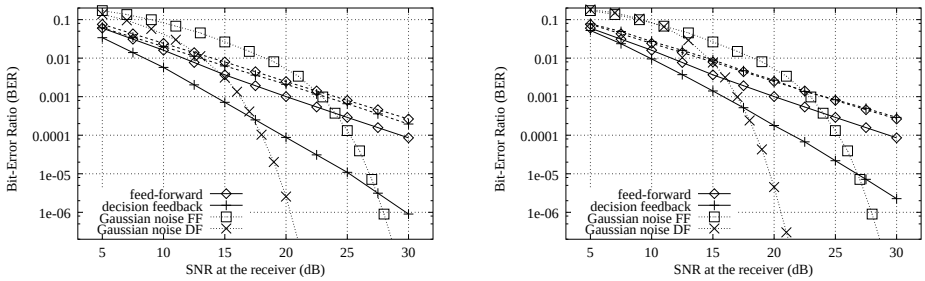
Fig. 6. Performance of the optimum (solid lines) and traditional (dashed lines) feed–forward Bayesian equalisers for a channel with 3 taps and a variety of values for α.

The performance of the the optimally designed MAP equaliser was recorded, along with that of the traditional Bayesian equaliser. The BER performance of both equalisers is plotted in fig. 6. It can be clearly seen that the optimum MAP equaliser outperforms the traditional Bayesian equaliser when the noise is non–Gaussian. In fact, the further the noise deviates from the Gaussian distribution, the more significant the performance degradation of the traditional Bayesian equaliser is.

For the simulations concerning the Bayesian DFE, α was set to 1. Again, both optimum and traditional equalisers were studied. The equaliser forward order was $M = 2$ and the decision lag $d = 1$, while the feedback order was $D = 2$. Fig. 7(a) shows the performance of the equalisers in this highly impulsive α–stable noise environment. For comparison, the BER graphs of the feed–forward and DF equalisers in Gaussian noise environment are given as well. In this experiment the correct transmitted data were fed in the feedback vector $\mathbf{x}_{\mathrm{D}}$ (for the DF equalisers). The results show that for 0.001 BER the performance benefit from the feed–forward optimum equaliser compared to the traditional one is 4.18 dB. The corresponding gain for the DF equaliser is 8.88 dB. For the next experiment (fig. 7(b)) the actual decisions of the DF equaliser were fed into the feedback vector $\mathbf{x}_{\mathrm{D}}$. As expected, the performance gain is slightly inferior (due to error propagation), but still considerable. For the DF equalisers this gain is 8.08 dB at 0.001 BER. That is, the use of the actual decision data results in a gain loss of 0.8 dB.

It is interesting to notice that the actual shape of the BER graphs for non–Gaussian stable noise is inherently different from the traditional graphs in Gaussian noise: the probability of error in a communication system is highly related to the probability of *exceedence* [6] $P_{x>\tau}$ of the underlying noise distri-

[6] The probability that the RV $\mathbf{x}$ *exceeds* (is greater than) a given τ.



(a) Correct data for the feed back. (b) Detected data for the feed back.

Fig. 7. Performance of the optimum (solid lines) and traditional (dashed lines) Bayesian DFE for $\alpha = 1$ ($M = 2$, $D = 2$, $d = 1$, $G = 4$).

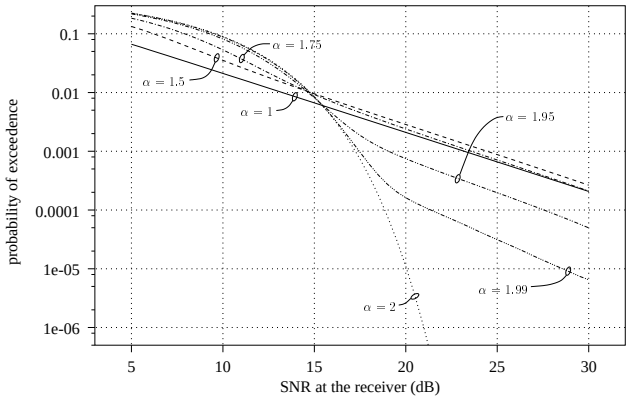


Fig. 8. Probability of exceedence of α–stable distribution for a variety of values for α ($\tau = .6$, $G = 4$)

bution. For the Gaussian case ($\alpha = 2$) the probability of exceedence $P_{x>\tau}(\alpha, \gamma)$ is

$$P_{x>\tau}(2, \gamma) = \frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\frac{\tau}{2\sqrt{\gamma}}\right) \tag{24}$$

and the Cauchy case ($\alpha = 1$)

$$P_{x>\tau}(1, \gamma) = \frac{1}{2} - \frac{\arctan\left(\frac{\tau}{\gamma}\right)}{\pi} \tag{25}$$

In fig. 8 we plot this probability as a function of the SNR at the receiver. Equation (22) was used in order to map the values of γ to the corresponding values of SNR. The similarity of fig. 8 with figures 6, 7(a) and 7(b) is clear.

5 Training the equaliser

The optimum Bayesian equaliser derived in section 3 is fully defined by two sets of parameters: a) the vector centers $\mathbf{c}_i$ and their associated signs s_i, and b) the parameters of the probability density function of the $S\alpha S$ noise, namely the characteristic exponent α and dispersion γ (see eq. (7)). This section addresses the problem of determining these parameters for the Bayesian equaliser in a non–Gaussian α–stable noise environment. For the estimation of the equaliser centers, the most popular approach first estimates the channel impulse response with a traditional linear adaptive algorithm and uses the resulting channel estimate to calculate the equaliser centers [11]. A number of adaptive algorithms have been proposed in the literature for channel estimation in non–Gaussian noise environments. A family of such algorithms is presented in section 5.1 along with some experimental results for the evaluation of their performance. The problem of estimating the stable parameters in a communications context is addressed in section 5.2.

5.1 Channel estimation in α–stable noise environments

Suppose that the channel estimation algorithm operates in supervised mode and let $\hat{\mathbf{h}}$ be the channel estimate. The output of filter $\hat{\mathbf{h}}$ is then $\hat{y}(k) = \hat{\mathbf{h}}^{\mathrm{T}}\mathbf{x}_{\mathrm{ch}}(k)$ and the estimation error [7] $e(k) = r(k) - \hat{y}(k)$. The optimisation criterion for linear regression estimation in Gaussian noise environments is usually the minimisation of a quadratic function of the estimation error. It is well known [5], though, that quadratic optimisation criteria are meaningless for non-Gaussian stable signals, because only moments of order less than α are finite.

In [5] the authors introduce the *least mean p–norm* (LMP) algorithm as a direct generalisation of least mean squares (LMS) [12] for α–stable environments. LMP is very similar to LMS and its basic recursion is

$$\hat{\mathbf{h}}(k+1) = \hat{\mathbf{h}}(k) + \mu \, \langle \xi(k) \rangle^{p-1} \, \mathbf{x}_{\mathrm{ch}}(k) \tag{26}$$

where $\langle s \rangle^p = \mathrm{sign}(s) \, |s|^p$. Here, $\mu > 0$ is the step–size parameter and

$$\xi(k) = y(k) - \hat{\mathbf{h}}^{\mathrm{T}}(k-1)\mathbf{x}(k) \tag{27}$$

is the *a priori* estimation error. The need, however, for faster converging algorithms often calls for least squares (LS) signal processing. The effort to

[7] According to fig. 4 we should use r_{L} but for the moment we ignore the limiter at the front end of the receiver.

provide robust versions of LS algorithms for non–Gaussian noise environments has received attention in the literature [3,13–16].

A heuristic method to deal with the outliers of non–Gaussian distributions was proposed in [16]. A traditional LS algorithm (e.g. RLS [12])is still used, but the channel estimate adaptation is inhibited when the received signal $r(k)$ is corrupted by a noise sample which can be characterised as an "outlier". In order to identify the outliers, order statistics is used. Consider a sorted vector containing the magnitude of the last Θ estimation error samples $\xi(k)$. If the current error sample lies among the top η largest past samples, the current observation is characterised as an "outlier". The experimental results suggest that OSRLS can achieve good performance in highly impulsive environments. Its main disadvantage, though, is that there is no way of determining the optimal values for the parameters η and Θ.

5.1.1 Recursive weighted least squares

The class of M–estimators is a robust version of the LS estimate, proposed by Huber [15]. Instead of minimising the sum of squared error, a less rapidly increasing function ρ of the error is used

$$\mathcal{J}_{\mathrm{M}} = \sum_{i=0}^{k} \lambda^{k-i} \rho(e(i)) \tag{28}$$

where λ $(0 < \lambda \leq 1)$ is an exponential weighting factor. Suppose that ρ has a derivative $\psi = \rho'$; then, the minimisation of eq. (28) implies

$$\sum_{i=0}^{k} \lambda^{k-i} \psi(e(i)) x(i-j) = 0, \quad j = 0, 1, \ldots N-1 \tag{29}$$

If we now define

$$\phi(x) = \frac{\psi(x)}{x}, \quad \text{and} \quad v(i) = \phi(e(i)) \tag{30}$$

equation (29) can then be rewritten as follows

$$\sum_{i=0}^{k} \lambda^{k-i} v(i) e(i) x(i-j) = 0, \quad j = 0, 1, \ldots N-1 \tag{31}$$

The sequence $v(i)$ assumes knowledge of the optimal weight vector $\hat{\mathbf{h}}$ at time k to generate the error sequence $e(i)$. As in [13,17], for the recursive approximation of the tap weight estimate $\hat{\mathbf{h}}$ the instantaneous *a priori* estimation error $\xi(i)$ is used to approximate $e(i)$. We can, therefore, generate the sequence

$$w(i) = \phi(\xi(i)) \tag{32}$$

in order to approximate $v(i)$. Elaborating as in [16], yields the *recursive weighted least squares* (RWLS) algorithm. Its basic recursion can be summarised as

$$\begin{aligned}\mathbf{K}(k) &= \frac{\lambda^{-1}\mathbf{P}(k-1)\mathbf{x}(k)}{w(k)^{-1}+\lambda^{-1}\mathbf{x}^{\mathrm{T}}(k)\mathbf{P}(k-1)\mathbf{x}(k)} \\ \hat{\mathbf{h}}(k) &= \hat{\mathbf{h}}(k-1)+\xi(k)\mathbf{K}(k) \\ \mathbf{P}(k) &= \lambda^{-1}\mathbf{P}(k-1)-\lambda^{-1}\mathbf{K}(k)\mathbf{x}^{\mathrm{T}}(k)\mathbf{P}(k-1)\end{aligned} \tag{33}$$

Traditional Recursive Least Squares (RLS):
The RWLS recursion is reduced to the traditional RLS algorithm if a quadratic penalty function is chosen, corresponding to

$$\rho_{\mathrm{LS}}(x)=\frac{x^2}{2}, \quad \psi_{\mathrm{LS}}(x)=x, \quad \phi_{\mathrm{LS}}(x)=1 \tag{34}$$

Recursive Maximum Likelihood (RML):
The likelihood function of the received vector $\mathbf{y}$ under the parameters $\hat{\mathbf{h}}$ is given by

$$\mathcal{L}_{\hat{\mathbf{h}}}(\mathbf{y};f_\alpha) \triangleq \log\prod_{i=0}^{k} f_\alpha(e(i)) = \sum_{i=0}^{k}\log f_\alpha(e(i)) \tag{35}$$

Therefore, the maximum likelihood estimate (ML) is given by RWLS when

$$\rho_{\mathrm{ML}}(x)=-\log f_\alpha(x), \quad \psi_{\mathrm{ML}}(x)=-\frac{f'_\alpha(x)}{f_\alpha(x)}, \quad \phi_{\mathrm{ML}}(x)=-\frac{f'_\alpha(x)}{x f_\alpha(x)} \tag{36}$$

Recursive Least p–norm (RLP):
As shown in [5], the *minimum dispersion* criterion is a natural and mathematically meaningful choice as a measure of optimality in stable signal processing. Consequently, the appropriate cost function would be $\mathcal{J}_{\mathrm{LP}}=\sum_{i=0}^{k}|e(i)|^p$, and a *recursive least p–norm* (RLP) algorithm is obtained by setting

$$\rho_{\mathrm{LP}}(x)=\frac{|x|^p}{p}, \quad \psi_{\mathrm{LP}}(x)=\langle x\rangle^{p-1}, \quad \phi_{\mathrm{LP}}(x)=|x|^{p-2} \tag{37}$$

Another approach to the least p–norm optimisation problem was taken by Byrd and Payne [18] in the form of the *iteratively re–weighted least squares* (IRLS) algorithm (also see [14,19]). The role of IRLS, however, in a communications signal processing context is questionable, because it requires infinitely growing storage memory and number of computations.

5.1.2 Bounding the weighting sequence

Equation (37) suggests that $w(i)$ is, in general, not bounded. That is, for infinitesimally small error the corresponding weight is infinitely large. Theoretical justifications [15,19,20] however, require a bounded weighting sequence. Huber [15], suggested that it is desirable to bound the sequence $w(i)$ for very small samples of the estimation error, as follows

$$w(i) = \begin{cases} |\xi(i)|^{p-2} \ , & |\xi(i)| > \omega \\ \omega^{p-2} \ , & |\xi(i)| \leq \omega \end{cases} \tag{38}$$

where ω is a small positive constant.

The LMP algorithm actually corresponds to an unbounded weighting sequence, since $\langle \xi(k) \rangle^{p-1} = |\xi(k)|^{p-2} \, \xi(k)$ (see eq. (26)). This results to a steep estimation error gradient close to zero, making LMP more sensitive to gradient noise in comparison with LMS. Replacing $|\xi(k)|^{p-2}$ with the bounded weighting sequence $w(i)$ from eq. (38) we obtain a stochastic gradient algorithm with less steep a gradient close to zero and therefore less misadjustment.

For a real–time channel estimation system we could also employ a time varying step–size parameter in order to speed–up the transient behaviour of LMP, such as $\mu(k) = (1 + c\nu^k) \, \mu_0$. Here ν is a constant controlling the speed of the transient of the step–size parameter ($0 < \nu < 1$) and $c > 0$. We can, therefore, summarise the recursion of a *modified* LMP (MLMP) as

$$\hat{\mathbf{h}}(k+1) = \hat{\mathbf{h}}(k) + (1 + c\nu^k) \, \mu_0 \, w(k) \, \xi(k) \, \mathbf{x}(k) \tag{39}$$

5.1.3 Experiments

Unfortunately, there is no convergence and stability analysis for LMP or RWLS. Nevertheless, The experimental data suggest that the algorithms converge efficiently and produce satisfactory estimates of the channel impulse response in impulsive noise environments. Our experiments have been carried out with a channel impulse response

$$\mathbf{h} = [\, .04 \quad -.05 \quad .07 \quad -.21 \quad -.5 \quad .72 \quad .36 \quad 0 \quad .21 \quad .03 \quad .07 \,]^{\mathrm{T}} \tag{40}$$

and noise parameters $\alpha = 1, \quad \gamma = .08$. The step–size parameter for LMP was $\mu = .004$, while for MLMP $\mu_0 = .004$, $\nu = 0.875$ and $c = 10$. For OSRLS, $\eta = 3$ and $\Theta = 9$. For RWLS λ was set to 0.98. Finally, ω was chosen 0.1 (for MLMP and RLP).

Figure 9 depicts the ensemble (over 400 Monte–Carlo runs) mean squared error (MSE) for algorithms LMP, MLMP, OSRLS, RML and RLP. IRLS is a block

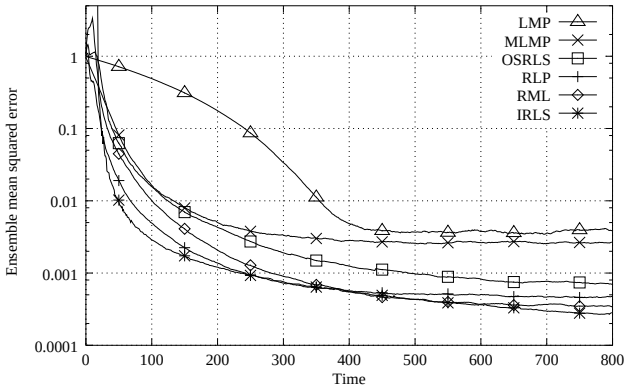


Fig. 9. The convergence of LMP, MLMP, OSRLS, RML and RLP for a channel with 11 taps

algorithm and can not be compared with this family of recursive algorithms in a direct manner. However, its ensemble convergence for the same constellation is also depicted in order to obtain a relative measure for the performance of the recursive algorithms, since IRLS offers the best known performance for the least p–norm optimisation problem.

Clearly, the LS type algorithms outperform the stochastic gradient ones. Furthermore, as expected, the convergence of MLMP is better than LMP in terms of both transient behaviour and steady state misadjustment. In fact, the transient convergence of MLMP is comparable with that of LS algorithms.

Among the recursive LS type algorithms, RML achieves, as expected, the best asymptotic performance. However, RML and RLP do not retain the defining characteristic of the traditional LS scheme, i.e., that the MSE continuously deminishes as $k \to \infty$. This behaviour, can be found in IRLS. On the contrary, the MSE for RML and RLP seems to reach an asymptotic infimum, a behaviour similar to the stochastic gradient algorithms. However, this infimum is significantly lower than LMP or MLMP. The transient behaviour of RLP is superior to RML, and actually comparable to that of IRLS. Finally, the performance of OSRLS is poorer than RML and RLP because, for the specific values of η and Θ, this algorithm discards about a third of the received samples.

In summary, IRLS offers the best known performance for least p–norm optimisation, but at an unaffordably high computational cost. Alternatively, there is a variety of recursive algorithms with reasonable complexity but compromised performance. Among these, the most suitable for channel estimation in a receiver are MLMP and RLP. They are both direct generalisations of the conventional LMS and RLS, respectively, with negligible extra computational requirements, providing robust performance in impulsive non–Gaussian environments.

5.2 Estimation of the noise parameters

Recall from section 2 that a $S\alpha S$ distribution is determined by three parameters: the characteristic exponent α, the dispersion γ, and the location parameter δ. In practice, an adaptive channel equaliser would be required to estimate the parameters of the noise from the actual received data. In most communication systems the noise is symmetric around zero, so we can assume that $\delta = 0$. For the estimation of α and γ, a variety of algorithms can be found in the literature [21]. These algorithms are based either on statistical quantiles [22],either on the sample characteristic function of the data [23–25],or on fractional lower order moments [26,27].

The quantile based techniques, although efficent in a statistical analysis environment, are not suitable for signal processing in a communications context. The characteristic function based scheme (Koutrouvelis' algorithm [25]),on the other hand, although computationally expensive, has been formulated as a linear regression problem. This characteristic and the fact that the implementation of this algorithm is straightforward are highly desirable in signal processing for communications. Furthermore, its estimates are consistent and unbiased [21].

In terms of efficient implementation and simplicity, however, the log FLOM algorithm, proposed by Ma and Nikias [26], is superior. This is a pure recursive algorithm, with minimal computational complexity and fairly simple implementation. Its main disadvantage, though, is that the convergence speed of the characteristic exponent estimate degrages for α close to 2. Nevertheless, the estimation of c is more robust, even though its computation involves the estimate for α.

According to our experiments, the sensitivity of the optimum Bayesian equaliser to the estimate of the characteristic exponent is small enough to accomodate the inaccuracy of this algorithm's estimates. This can be clearly seen in fig. 10. Hence, the adoption of the algorithm proposed by Ma and Nikias is considered adequate for estimating the stable parameters in a Bayesian equaliser. Therefore, this algorithm will be used in the rest of this paper in order to evaluate the performance of the adaptive Bayesian equaliser in an α–stable noise environment.

6 Performance analysis of the adaptive equaliser

This section discusses the performance of a complete adaptive equaliser, like the one shown in fig. 4. A set of simulation experiments were carried out in

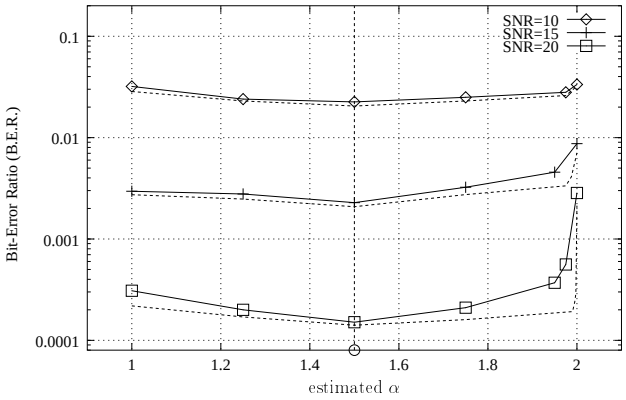


Fig. 10. Robustness of the adaptive (RLP) Bayesian DFE equaliser (solid lines) with respect to the estimated characteristic exponent α for actual $\alpha = 1.5$ (the dashed lines correspond to perfect channel knowledge).

order to investigate the performance of the adaptive Bayesian DFE. The transmitted data were organised in frames of 128 bits with the first 32 bits serving as pilot data. The symbol rate is assumed to be 300 KHz. Both stationary and Rayleigh time–varying scenarios were simulated. For the latter, the taps of the non–stationary channel were correlated Rayleigh RV's multiplied by the appropriate tap root–mean–power (RMP).

The Rayleigh RV's were generated using the deterministic approach proposed by Rice [28,29]. In this scheme, the real and imaginary parts of the complex coloured Gaussian RV are formed as sums of sinusoids. The statistical properties of this scheme are extracted by Pätzold et. al. in [30]. For the shape of the Doppler power spectral density (PSD) of the complex Gaussian noise process we adopt the Jakes PSD [31] for mobile fading channel models.

Note that for the non–stationary channel scenario, the signal–to–noise ratio at the receiver is defined [11] as

$$\mathrm{SNR}_{\mathrm{rcv}} = \frac{\lim\limits_{k\to\infty} \frac{1}{k} \sum\limits_{i=0}^{k-1} E_x \left[|r_{\mathrm{L}}(k)|^2 \right]}{E_n \left[|\hat{n}(k)|^2 \right]} \tag{41}$$

where $E_p[\cdot]$ denotes the expectation operator with respect to the random process p.

For the channel estimator two algorithms were used, MLMP and RLP. The adaptation of the channel estimate taps takes place in both the training period and the data transmission period. For the former the known training sequence is used and for the latter the actual decisions of the equaliser are fed into the algorithm (decision–directed adaptation). The forward order of the equaliser was set to $M = 3$. For the order of the feed–back section (D) and the decision

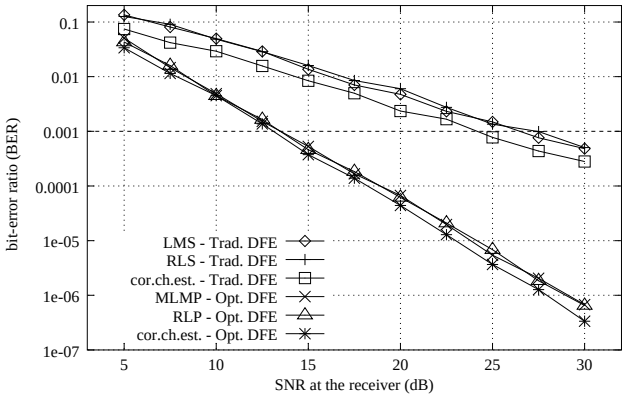


Fig. 11. Performance of the adaptive Bayesian DFE for $\alpha = 1$. The channel is stationary with 3 taps.

lag (d) the guidelines in [11] were used ($D = M - 1$ and $d = N - 1$). The α–stable parameters were estimated using the log FLOM algorithm.

Experiment 1: *Stationary channel*
The stationary channel consists of 3 taps

$$\mathbf{h} = [0.3482 \quad 0.8704 \quad 0.3482]^{\mathrm{T}} \tag{42}$$

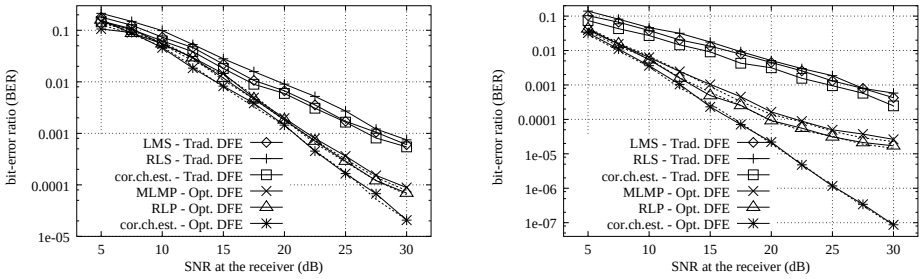
and the noise characteristic exponent is $\alpha = 1$. For this experiment, the correct stable parameters were provided to the equalisers. The dynamic range of the receiver was $G = 4$. Figure 11 depicts the performance of both optimum and traditional adaptive Bayesian DFE's. The performance of the equalisers with a perfect channel estimation is given as well.

The optimum adaptive DFE has a performance which is very close to the optimal and it seems that both MLMP and RLP perform equally well in a stationary environment. On the other hand, the traditional adaptive DFE suffers a significant performance degradation in this highly impulsive noise environment. For example the 1/1000 performance target is achieved by the optimum equaliser at 13.86 dB, while the same target is reached by the traditional equaliser at 27.42 dB (for RLP and RLS channel estimators, respectively) resulting to a benefit of 13.56 dB.

Experiment 2: *Rayleigh fading channel*
The Rayleigh fading channel consists of 3 taps with RMP's

$$\mathbf{h} = [0.3482 \quad 0.8704 \quad 0.3482]^{\mathrm{T}} \tag{43}$$

This experiment was carried out for both true and estimated α–stable parameters. The dynamic range of the receiver was $G = 6$. The performance of both the optimum and traditional adaptive Bayesian DFE was recorded in a



(a) $\alpha = 1.5$ (b) $\alpha = 1$

Fig. 12. Performance of the adaptive Bayesian DFE with noise parameters estimates (solid lines) and actual noise paramters (dashed lines) for the non–stationary channel.

noise environment with characteristic exponent $\alpha = 1.5$ and $\alpha = 1$ (fig. 12(a) and fig. 12(b), respectively). The results with estimated stable parameters are depicted in solid lines, while the dashed lines correspond to true parameters. For comparison, the performance of the equalisers with a perfect channel estimation is given as well.

As expected, there is a definite performance loss of the adaptive DFE in comparison with the non-stationary scenario, due to the limited tracking ability of the channel estimation algorithms and the fading characteristics of the channel. However, the performance advantage of the optimum adaptive Bayesian DFE is still significant, compared to the bit–error ratio of the corresponding traditional DFE (i.e. designed under the Gaussian assumption). For the case when true stable parameters are used, this benefit for $\alpha = 1.5$ is 6.91 dB at a BER of 0.001. For $\alpha = 1$ the performance gain is 12.98 dB for the same BER of 0.001.

These results also indicate that the utilisation of estimates rather than the actual values for the noise parameters does not practically compromise the performance of the equalisers. This suggests that the dominant factor affecting the performance of the adaptive equaliser is the design of the Bayesian (MAP) detector and the channel estimation algorithms.

The tracking performance of RLP in this fast changing environment is marginally better than MLMP, especially for the highly impulsive noise environment $\alpha = 1$ (fig. 12(b)). This is exactly the opposite situation to the Gaussian noise environment [11], where the stochastic gradient algorithm achieves better tracking of the channel than the least squares approach. This dissimilarity should be attributed to the noise statistics and the actual cost function of the channel estimation algorithms.

The principal consequence of a non-quadratic cost function is a noticeable

deterioration of the tracking ability of the adaptive algorithms as the characteristic exponent moves from 2 to 1. Recall from section 5.1.3 that in a highly impulsive noise environment, the weighting sequence $w(i)$ (eq. (32)) suppresses the samples with large estimation error, because they are likely to be the result of noise impulses. But, when the channel is non-stationary, large residuals can often arise as a result of the discrepancy between the channel estimate and the actual channel impulse response. The suppression of these residuals effectively decelerates the tracking ability of the channel estimation algorithms.

7 Practical approximations for stable distributions

Unfortunately, the performance benefit of the proposed equaliser comes at the expense of high computational load. Recall from section 2 that closed form α–stable densities only exist for $\alpha = 2$ (Gaussian) and $\alpha = 1$ (Cauchy). In all other cases, numerical approximation of the stable distribution is required, making the use of stable distributions in real time systems unaffordable. However, in MAP applications, only the actual shape of the decision boundary is important for the performance of the equaliser, which means that an approximation to the stable density may be used. Here, we propose a linear interpolation between the Gaussian and the Cauchy distribution as an approximation to the stable density for $1 < \alpha < 2$, i.e.

$$\hat{f}_\alpha(s) = (1-p)\ f_1(s) + p\ f_2(s) \tag{44}$$

where p is an increasingly monotonic function of α ($0 \leq p \leq 1$ for $1 \leq \alpha \leq 2$). $f_2(x)$ and $f_1(x)$ are the Gaussian and Cauchy distributions, respectively [8] .

Assuming that the dynamic range of the receiver can accomodate all scalar centers without distortion ($G > \bar{c}_i,\ \forall i$), it is only required that the symmetric noise pdf is approximated within the range $[0, 2G]$. Furthermore, the shape of the noise distribution close to the origin does not affect the optimum decision boundary. Therefore, the approximation range can further be reduced to $[H, 2G]$, where $0 < H < 2G$. The optimum value for p with respect to α can then be derived by a least squares optimisation of the form

$$p_{\text{opt}}(\alpha) = \arg \min_{0<p<1} \int_H^{2G} \left| f_\alpha(s) - (1-p) f_1(s) - p f_2(s) \right|^2 ds \tag{45}$$

We have numerically solved eq. (45) for a number of values of α with $G = 4$, $H = 2$, and $\gamma = 0.01$. The resulting set of optimum values for p is depicted

[8] A RV generated by time multiplexing a Gaussian RV with probability p and a Cauchy RV with probability $(1-p)$, actually has a pdf given by eq. (44).

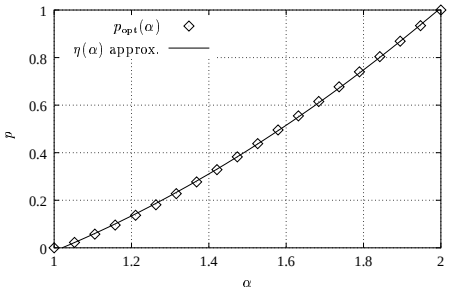


Fig. 13. The actual and approximated p_{opt} as a function of α.

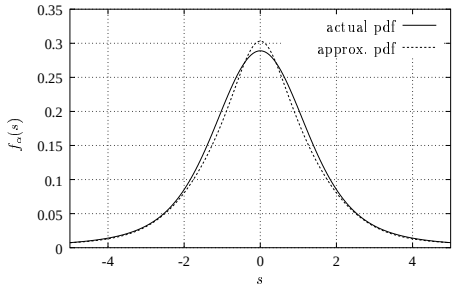


Fig. 14. The actual and approximated α–stable pdf for $\alpha = 1.5$ ($\gamma = 1$).

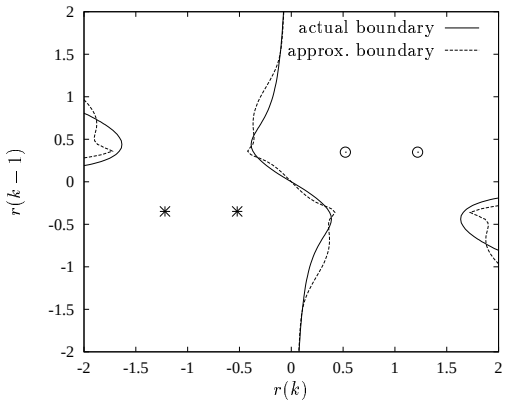


Fig. 15. Actual and approximated decision boundary of the Bayesian DFE with channel $H(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$ for $\alpha = 1.5$ ($M = 2, D = 2, d = 1$).

in fig. 13. It would be desirable, however, to approximate $p_{\text{opt}}(\alpha)$ with a more simple formula. We can, for example, apply second degree polynomial fitting on the set of optimum values for p obtained from eq. (45) to produce a relation $p = \eta(\alpha)$. This relation has been found to be (see fig. 13)

$$\eta(\alpha) = 0.3521\ \alpha^2 - 0.0329\ \alpha - 0.3333 \tag{46}$$

Figure 14 shows the actual and the approximated pdf for $\alpha = 1.5$. Furthermore, as fig. 15 shows, the approximated pdf produces a decision boundary which preserves the features of the optimum boundary. Fig. 16 depicts the BER performance of the Bayesian DFE in a stationary channel with 3 taps with the approximated pdf (solid lines) and the true pdf (dashed lines). These results suggest that the approximation results to a performance loss of less than 2 dB for $\alpha = 1.5$. For $\alpha = 1.95$ the performance loss is indistinguishable.

In section 4, we analytically derived the variance of the noise estimate $\hat{n}$ (eq. (21)) for $\alpha = 1$ and $\alpha = 2$. In all other cases, it is only possible to measure this variance experimentally. Our experiments, however, suggest that as α moves

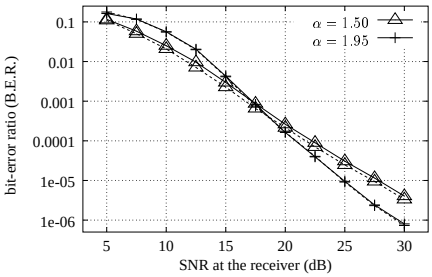


Fig. 16. Performance of the Bayesian DFE with approximated α–stable distribution ($G = 4$).

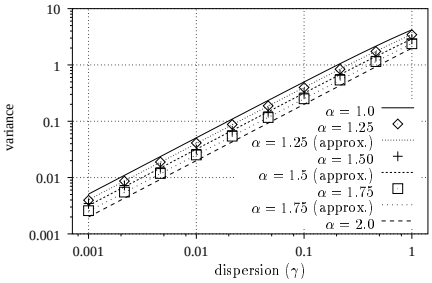


Fig. 17. Actual and approximated variance with respect to the noise dispersion γ ($G = 4$).

from 1 to 2 the variance v_α of the noise estimate $\hat{n}$ moves in a linear way (in the log domain) from v_1 to v_2 (the calculated variances for $\alpha = 1$ and $\alpha = 2$, respectively). Therefore, a reasonable approximation should be

$$\hat{v}_{\hat{n}}(\alpha, \gamma, G) = v_{\hat{n}}(1, \gamma, G)^{2-\alpha}\ v_{\hat{n}}(2, \gamma, G)^{\alpha-1} \tag{47}$$

Figure 17 depicts the experimental (true) $v_{\hat{n}}$ and approximated $\hat{v}_{\hat{n}}$ with respect to the noise dispersion γ for different values of α and $G = 4$. Figure 18, on the other hand, shows $v_{\hat{n}}$ and $\hat{v}_{\hat{n}}$ as a function of the limiting level G for different values of α. These graphs show that the approximation of eq. (47) is sufficiently satisfactory for a wide range of the limiting level G and noise dispersion γ.

In a similar way as eq. (47), we can obtain a good approximation of the appropriate dispersion γ for a given $\mathrm{SNR}_{\mathrm{rcv}}$ when α is not equal to 1 or 2. More precisely, this approximation is

$$\hat{\gamma}_\alpha = \gamma_1^{2-\alpha}\ \gamma_2^{\alpha-1} \tag{48}$$

where γ_α ($\alpha = 1, 2$) is the solution of equation

$$v_{\hat{n}}(\alpha, \gamma_\alpha, G) = \frac{\Delta V^2}{2 \cdot 10^{\mathrm{SNR}_{\mathrm{rcv}}/10}} \tag{49}$$

Figure 19 shows that this approximation is reliable for a wide range of SNR's.

8 Conclusion

The optimum adaptive Bayesian DFE for α–stable noise environments was presented and its performance in a variety of channel scenarios, stationary and Rayleigh fading, was investigated. For the adaptation of the equaliser

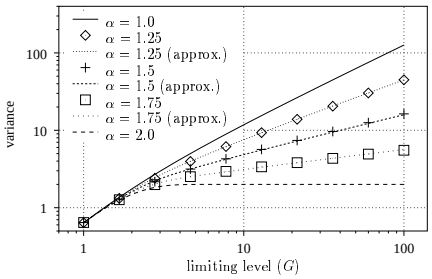


Fig. 18. Actual and approximated variance with respect to the limiting level G ($\gamma = 1$).

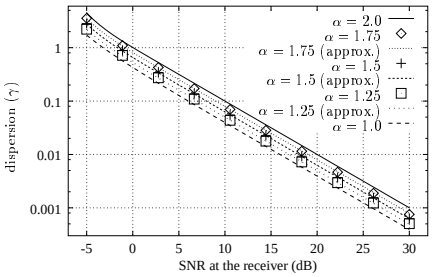


Fig. 19. Actual and approximated noise dispersion γ ($G = 4$).

centers a family of generalised channel estimation algorithms was used. In order to quantitatively assess systems in such infinite power noise environments, a new analytical framework was proposed as well. Compared with a conventional adaptive Bayesian DFE designed under the Gaussian assumption, the proposed adaptive equaliser exhibits a significant performance advantage. Unfortunately, the computational overhead for the computation of the α–stable density may not always be affordable. However, certain practical approximations were presented offering near–optimum performance with negligible complexity surcharge.

References

[1] K. Abend and B.D. Fritchman. Statistical detection for communication channels with intersymbol interference. *Proceedings of the IEEE*, 58:779–785, 1970.

[2] B. Mulgrew. Applying radial basis functions. *IEEE Sig. Proc. Magazine*, pages 50–65, March 1996.

[3] X. Wang and H.V. Poor. Robust multi–user detection in non–Gaussian channels. *IEEE Transactions on Signal Processing*, 47(2):289–305, February 1999.

[4] D. Middleton. Non–Gaussian noise models in signal processing for telecommunications: New methods and results for class A and class B noise models. *IEEE Transactions on Information Theory*, 45(4):1129–1149, May 1999.

[5] M. Shao and C. L. Nikias. Signal processing with fractional lower order moments: Stable processes and their applications. *Proceedings of the IEEE*, 81:986–1009, July 1993.

[6] W. Feller. *An introduction to probability theory and its applications.* Wiley, New York, 1971.

[7] G. Samorodnitsky and M.S. Taqqu. *Stable non–Gaussian random processes.* Chapman & Hall, 1994.

[8] A. T. Georgiadis and B. Mulgrew. A MAP equaliser for impulsive noise environments. *First IMA International Conference on Mathematics in Communications*, December 1998.

[9] A.P. Clark, L.H. Lee, and R.S. Marshall. Developments of the conventional nonlinear equaliser. *IEE Proccedings*, 129(2):85–94, 1982.

[10] D. Williamson, R. A. Kennedy, and G. W. Pulford. Block decision feedback equalization. *IEEE Transactions on Communications*, 40:255–264, February 1992.

[11] S. Chen, S. McLaughlin, B. Mulgrew, and P. Grant. Adaptive bayesian decision feedback equaliser for dispersive mobile radio channels. *IEEE Transactions on Communications*, 43(5):1937–1945, May 1995.

[12] S. Haykin. *Adaptive Filter Theory.* Prentice–Hall, 3rd edition, 1997.

[13] H. Dai and N.K. Sinha. Robust recursive least–squares method with modified weights for bilinear system identification. *Proceedings of the IEE*, 136(3):122–126, May 1989.

[14] E. Kuruoğlu, W. Fitzgerald, and P. Rayner. Non–linear autoregressive modeling of non–Gaussian signals using L_p–norm techniques. *International Conference on Acoustics, Speech and Signal Processing*, 3:3533–3536, 1997.

[15] P.J. Huber. *Robust Statistics.* Wiley, New York, 1981.

[16] A. T. Georgiadis and B. Mulgrew. A family of recursive algorithms for channel identification in alpha–stable noise. *5th Bayona workshop on emerging technologies in telecommunications*, pages 153–157, September 1999.

[17] S.C. Puthenpura, N.K. Sinha, and O.P. Vidal. Application of M–estimation in robust recursive system identification. *IFAC Symp. Stocastic Control*, pages 23–30, 1985.

[18] R H Byrd and D A Payne. Convergence of the iteratively reweighted least squares algorithm for robust regression. Technical report 313, The Johns Hopkins Univ. Baltimore, MD, June 1979.

[19] R. Yarlagadda, J.B. Bednar, and T. Watt. Fast algorithms for l_p deconvolution. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 33:174–182, 1985.

[20] L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification.* MIT Press, 1983.

[21] V. Akgiray and C.G. Lamoureux. Estimation of stable–law parameters: A comparateive study. *J. Business and Economic Statistics*, 7:85–93, January 1989. a–stable parameters estimation.

[22] E.F. Fama and R. Roll. Parameter estimates for symmetric stable distributions. *Journal of the American Statistical Association*, 66:331–338, June 1971. a–stable parameters estimation.

[23] S.J. Press. Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association*, 67:842–846, December 1972. a–stable parameters estimation.

[24] A.S. Paulson, E.W. Holcomb, and R. Leitch. The estimation of the parameters of the stable laws. *Biometrika*, 62:163–170, 1975. a–stable parameters estimation.

[25] I.A. Koutrouvelis. Regression–type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, 75:918–928, December 1980. a–stable parameters estimation.

[26] X Ma and C L Nikias. Parameter estimation and blind channel identification in impulsive signal environments. *IEEE Transactions on Signal Processing*, 43:2884–2897, December 1995.

[27] G A Tsihrintzis and C L Nikias. Fast estimation of the parameters of alpha–stable impulsive interference using asymptotic exterme value theory. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pages 1840–1843, 1995.

[28] S O Rice. Mathematical analysis of random noise. *Bell Systems Technical Journal*, 23:282–332, July 1944.

[29] S O Rice. Mathematical analysis of random noise. *Bell Systems Technical Journal*, 24:46–156, January 1945.

[30] M Pätzold, U Killat, F Laue, and Y Li. On the statistical properties of deterministic simulation models for mobile fading channels. *IEEE Transactions on Vehicular Technology*, 47(1):254–269, February 1998.

[31] W C Jakes, editor. *Microwave Mobile Communications*. IEEE Press, New York, 1993.